Engineering Thermostable Fungal Cellobiohydrolases
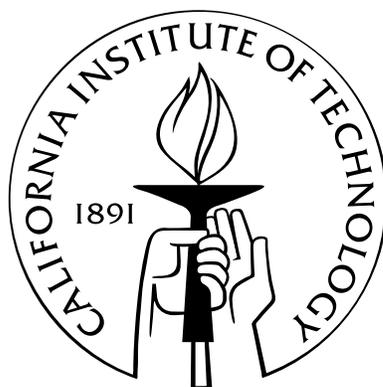
Thesis by

Indira Wu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2013

(Defended March 26, 2013)

Acknowledgements

Caltech has been an inspiring place to study and conduct research. The people I met during my time here all share so much passion for their crafts and vision for the future. It is truly an honor to be amongst so many driven and successful individuals. First and foremost, I want to express my deepest gratitude to my thesis advisor, Frances Arnold, for her relentless drive to better the world with engineered enzymes and her patience to guide me over the bumps and road blocks in my project. My stubbornness in believing in brute-force attempts at overcoming walls cannot possibly be easy to deal with. I want to thank my thesis committee: Doug Rees, Steve Mayo, and Dave Tirrell, for their generosity in sharing their expertise and advice. I am also grateful to Niles Pierce for constantly improving the Bioengineering option and to my fellow BE classmates for being great companions and support throughout the process.

I also cannot ask for a better lab to work in than the Arnold lab. The many graduate students and post-docs have not only made the lab an expertise-filled place but also a truly enjoyable environment to work in. Over the years, Florence Mingardon, Andrea Rentmeister, Russ Komor, Mary Farrow, Matt Smith, Devin Trudeau, Claire Bedbrook, and Thomas Heel have made working with cellulases enjoyable, and I am really grateful for that. I also want to thank Rich Heins who is an amazing colleague I met at JBEI and the sole reason that the ionic liquid work was done in a short span of 10 days. My thanks also go to Eric Brustad and Chris Snow for teaching me crystallography, Phil Romero for helping me with consensus sequence determination, and Kersten Rabe for all the discussion around FoldX. Ryan Lauchli, John McIntosh, and Pedro Coelho are really fun officemates to have, and I do not know what I would do without all their subgroup meetings and overseas Skype calls in the office. Last but definitely not least, I also want to thank Sabine Brinkmann-Chen for

keeping the chaos in the lab at a minimum (as much as it *can* be with 30 people in the lab) and for graciously offering to proofread my thesis.

I am indebted to my parents, Raymond and Frances, and my brother James and SIL Christy, who are the most wonderful role models in life and in their professional careers. Without their unconditional support, I would not be at Caltech today. I also want to thank my husband Michael for being the most amazing man I have ever met: generous, caring, and considerate. Without him being the constant positive force in my life, I would not have been able to finish my thesis. He is also my most invaluable technical support, and for that I am eternally grateful. I am also thankful for my friends Joyce, Rushan, and Heikki, who have kept me tethered to the world outside of science and are just such a joy to be around.

Abstract

Meeting the world's growing energy demands while protecting our fragile environment is a challenging issue. Second generation biofuels are liquid fuels like long-chain alcohol produced from lignocellulosic biomass. To reduce the cost of biofuel production, we engineered fungal family 6 cellobiohydrolases (Cel6A) for enhanced thermostability using random mutagenesis and recombination of beneficial mutations. During long-time hydrolysis, engineered thermostable cellulases hydrolyze more sugars than wild-type Cel6A as single enzymes and binary mixtures at their respective optimum temperatures. Engineered thermostable cellulases exhibit synergy in binary mixtures similar to wild-type cellulases, demonstrating the utility of engineering individual cellulases to produce novel thermostable mixtures. Crystal structures of the engineered thermostable cellulases indicate that the stabilization comes from improved hydrophobic interactions and restricted loop conformations by proline substitutions. At high temperature, free cysteines contribute to irreversible thermal inactivation in engineered thermostable Cel6A and wild-type Cel6A. The mechanism of thermal inactivation in this cellulase family is consistent with disulfide bond degradation and thiol-disulfide exchange. Enhancing the thermostability of Cel6A also increases tolerance to pretreatment chemicals, demonstrated by the strong correlation between thermostability and tolerance to 1-ethyl-3-methylimidazolium acetate. Several semi-rational protein engineering approaches — on the basis of consensus sequence analysis, proline stabilization, FoldX energy calculation, and high B-factors — were evaluated to further enhance the thermostability of Cel6A.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Meeting the world's growing energy demands while protecting our fragile environment is a challenging issue. The total energy consumption in the US was >87 quadrillion BTU in 2012[1]; 82% of the consumed energy was from fossil fuels like petroleum, coal, and natural gas, and only 9% was from renewable sources like solar, wind, hydroelectric power, and biomass (Figure 1.1) [1]. Despite concerns that the emission of carbon dioxide and other pollutants from fossil fuel combustion contributes to global warming, the future of alternative energy is unclear [2]. Energy policies and events like the 2011 Fukushima nuclear meltdown or the discovery of shale oil on US soil can quickly change the energy landscape and the public opinion on alternative energy [3]. Nonetheless, long-term use of fossil fuels is not sustainable, and developing near-term renewable energy like biofuels is important for transitioning into long-term renewable energy like solar fuels, whose feasibility depends on major technology breakthrough [4, 5].

One of near-term renewable energy options is biofuel. Second generation biofuels are liquid fuels like long-chain alcohols produced from plant biomass. One of the key biopolymers for biofuel production is cellulose [6]. Along with lignin and hemicellulose, cellulose is the main structural component of most plant cell walls, making it the most abundant biopolymer in the world. Cellulose is a polysaccharide composed of glucosyl units linked by β-1,4 glycosidic bonds. The chain ends can be characterized as the reducing end by the potential aldehyde group on the C1 atom or the

---

[1]Energy consumption data was collected from January to November of 2012.
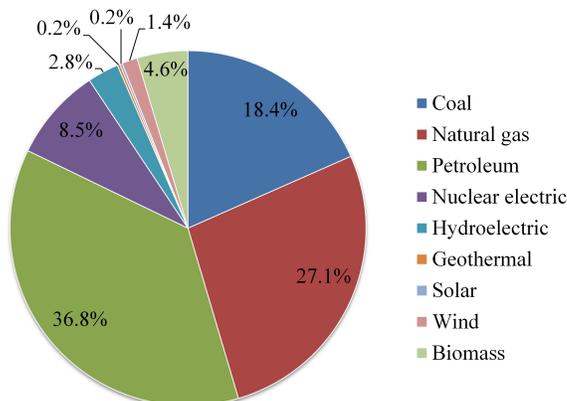
Source: US Energy Information Administration, Feb 2013

Figure 1.1: Total US energy consumption from January to November 2012 displayed by the energy source in percentage. The pie graph was generated using data from the Monthly Energy Review published by the US Energy Information Administration [1].

non-reducing end by the terminal hydroxyl group. The β-1,4 linkage ensures that the subunits rotate 180° every two glucose subunits, resulting in straight chains that can be bundled together tightly [7]. X-ray diffraction and nuclear magnetic resonance studies have shown that cellulose chains form intramolecular and intermolecular hydrogen bonds between the hydroxyl groups and the oxygen in the pyranose ring. This extensive hydrogen bond network produces highly crystalline elementary fibrils with strong tensile strength and low accessibility [8, 9]. In the absence of cellulases, cellulose is a highly recalcitrant material, with a half-life of over four million years at 25°C [10].

To utilize biomass for the production of fuels and higher-value chemicals, cellulose needs to be hydrolyzed into glucose for microbial fermentation. Concentrated hydrochloric and sulfuric acids (40 – 90%) have been used to hydrolyze cellulose and hemicellulose from plant matter into monomeric sugars. However, the high cost of the acids and the difficulty in acid recovery make the approach economically unattractive [11]. Fungal cellulases are an alternative solution for hydrolysis and have been used in industrial applications from cotton softening to pulp refining. For conversion of lignocellulosic biomass to fermentable sugar, physical and chemical pretreatments need to be applied prior to enzymatic degradation. Different pretreatment methods, structure-function relationship of fungal cellulases, as well as engineering challenges are reviewed below.

1.2 Pretreatments of lignocellulose

1.2.1 Physical pretreatments

To increase the accessibility of lignocellulose for subsequent enzymatic degradation, different physical and chemical processes have been applied to pretreat the substrate. Physical pretreatments involve different methods to reduce the particle size of the plant substrate, including chipping, milling, and grinding [12, 11]. The reduced particle size allows the subsequent chemical pretreatments to proceed more efficiently due to an increased surface area to mass ratio [13]. However, the energy consumption of milling varies greatly depending on the milling condition and can take up to 40% of the energy from the ethanol produced. To reduce the energy consumption, it might be preferable to integrate milling and chemical pretreatment into a multi-step process, i.e. subject wood chips to steam explosion before they are subjected to additional size reduction process [14].

1.2.2 Chemical pretreatments

Chemical pretreatments can be roughly divided into four categories: acid pretreatment, base pretreatment, steam explosion, and organic solvent fractionation. Acid pretreatments involve incubating lignocellulosic substrates with dilute $(0.5 - 2\%)$ sulfuric acids at $140 - 220°C$ [15]. Depending on the pretreatment temperature and sulfuric acid concentration, up to 90% of xylose can be released during pretreatment, leaving cellulose-rich solids after washing and detoxification [16]. But degradation products like furfurals or 5-hydroxymethylfurfural are often produced during acid preatreatments and can prohibit subsequent microbial fermentations [11]. Base pretreatments involve treating the substrate with ammonia, calcium oxide, or other base as catalysts [17]. These base catalysts increase substrate accessibility by removing acetyl groups from hemicellulose and lignin [13].

Steam explosion involves incubating the substrate at high temperature (up to $260°C$) and high pressure for a few seconds before rapidly dropping the pressure and allowing the wood fiber to expand [11]. For ammonia fiber expansion/explosion (AFEX), the substrate is pre-wetted with dry ammonia and incubated at $130°C$ and 650 psi for up to an hour before the pressure is purged. The

advantage of AFEX is that ammonia is relatively easy to separate from the substrate compared to other pretreatment chemicals. However, hemicellulases are required in addition to cellulases for the hydrolysis of AFEX-pretreated substrate since AFEX does not solubilize hemicellulose [13].

Organic solvent pretreatments include methods that utilize organosolv and ionic liquids. For organosolv process, mixtures of solvents (e.g. ethanol) and acids (e.g. sulfuric acids) are incubated with the substrate at $100 - 250°$C to solubilize hemicellulose and separate lignin from cellulose [11]. Ionic pretreatment involves solubilizing cellulose and lignin with low melting point salts and is discussed in detail in Chapter 4. Regardless of the chemical pretreatment methods, operating reactors and handling hazardous chemicals at high temperature and/or pressures are typically required. Detailed economic analysis and Green Chemistry assessments are necessary to identify the pretreatments most suitable for commercialization [18].

1.3   Fungal cellulases

1.3.1   Cellulases secreted by *Hypocrea jecorina*

Despite the recalcitrance of cellulose, Nature has provided several enzyme solutions to hydrolyzing cellulose into monomeric sugars that are utilized by microorganisms as a source of carbon and energy. The mesophilic fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*) secretes an array of cellulases that work synergistically to degrade cellulose to smaller oligomers and eventually to glucose. Cellulases secreted by *H. jecorina* include at least five endoglucanases (EGI-V), two cellobiohydrolases (CBHI belonging to glycoside hydrolase family 7 and CBHII belonging to glycoside hydrolase family 6), two β-glucosidases, and numerous hemicellulases. In particular, endoglucanases randomly attack β-1,4 glycosidic bonds in the amorphous regions of the cellulose and increase the available chain ends for cellobiohydrolases. Cellobiohydrolases incrementally shorten the chain length and include enzymes that act on the reducing ends (CBHI or Cel7A) and the non-reducing ends (CBHII or Cel6A) of cellulose. β-glucosidases hydrolyze soluble oligomers into glucose and reduce product inhibition for endoglucanases and cellobiohydrolases [19, 9].

Side view   Top view
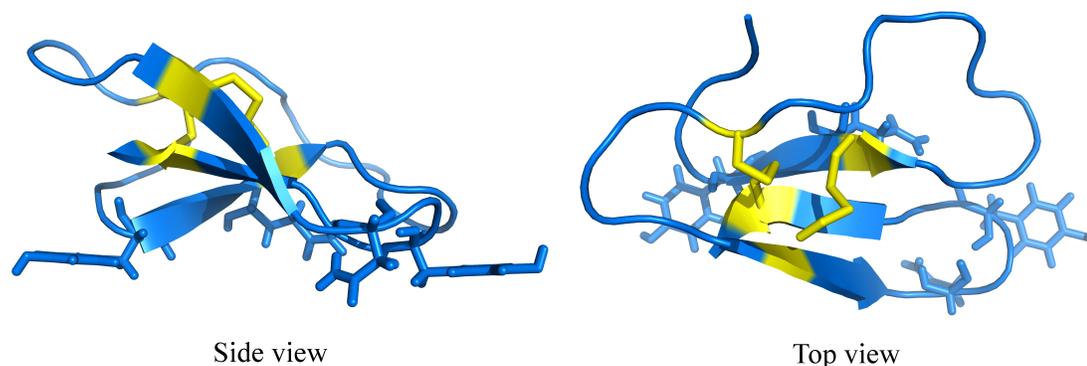
Figure 1.2: NMR structure of family 1 cellulose binding module from *H. jecorina* (PDB 1CBH, [23]). The disulfide bridges are shown in yellow, and the conserved residues are shown in sticks.

Cellobiohydrolases Cel7A, Cel6A, and endoglucanase Cel5A (EGII) comprise $60 \pm 5\%$, $20 \pm 6\%$, and $12 \pm 3\%$ of total protein secreted by *H. jecorina* [7]. All three cellulases share the same feature of a cellulose-binding domain and a catalytic domain connected by a highly O-glycosylated peptide linker [9]. Both the Cel6A and Cel7A catalytic domains feature tunnel-shaped structures formed by loops that are stabilized by disulfide bridges. In the Cel7A catalytic domain, two large antiparallel, concaved β sheets form a β sandwich and together with long loops stabilized by nine disulfide bonds form a $\sim$40 Å tunnel [20]. In the Cel6A catalytic domain, two extensive loops stabilized by two disulfide bonds and an α/β-barrel structure form a $\sim$20 Å tunnel [21]. Cellobiose is the primary product of cellulose hydrolysis by cellobiohydrolase Cel6A and Cel7A. However, it is worth noting that the division of endo- and exo-acting cellulases is not absolute. Cel6A from *Humicola insolens* has been reported to exhibit some endoglucanase activity, as have other cellobiohydrolases [22].

1.3.2 Cellulose binding module

The cellulose binding modules (CBM) from *H. jecorina* Cel6A and from most fungal cellulases belong to family 1 CBM that bind specifically to crystalline cellulose [24, 25]. CBM1 contains roughly 36 residues and two to three disulfide bonds [26]. Previous studies on substrate binding and processivity suggest that the flat binding face of CBM1 is composed of Y5/W5, Q7, N29, Y31/W31, and Y32 (Figure 1.2), and these residues are highly conserved among CBM1 [24]. CBM1 preferentially binds
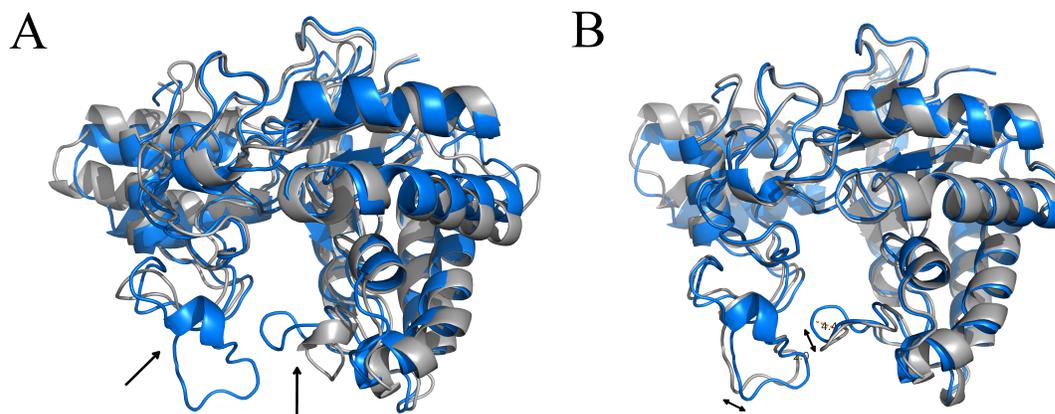
Figure 1.3: Structure alignments of Cel6A cellobiohydrolases from *Humicola insolens*. A) Comparison between Cel6A cellobiohydrolase (blue, PDB 1OCN chain A, [30]) and Cel6B endoglucanase (grey, PDB 1DYS chain A, [31]). Arrows indicate loop bending and truncation between a cellobiohydrolase and an endoglucanase. B) Comparison between Cel6A with substrate bound (blue, PDB 2BVW chain A, [32]) and without (grey, PDB 1BVW, [33]). Arrows indicate movements of the active site-forming loops between open and closed conformations.

to the hydrophobic surface over the hydrophilic surface of the cellulose microfibrils, and binding is dominated by the hydrophobic interactions between the aromatic side chains and the pyranose rings [26, 27]. However, since lignin is composed of phenolic compounds and consequently very hydrophobic, CBM1 also binds to lignin non-specifically, which reduces the productivity of cellulases [28, 29].

1.3.3 Glycoside hydrolase family 6

Glycoside hydrolase family 6 (GH6 or Cel6) includes many fungal cellobiohydrolases and endoglucanases that are important in the enzymatic degradation of cellulose to fermentable sugars [34, 35]. Structures of nine Cel6 cellulases and their variants have been determined, and they share a distorted $(\beta/\alpha)_7$ barrel fold with either a tunnel-shaped or cleft-shaped active site composed primarily of surface loops (Figure 1.3A) [36, 31]. The different preferences in substrate sites between cellobiohydrolases and endoglucanases stem from the shape of the active site. The open cleft-shaped active site allows endoglucanases to slide along the cellulose chain easily, while the closed tunnel-shaped active site allows cellobiohydrolases to thread the cellulose chain through and incrementally cleave glycosidic bonds. Interestingly, structure information of Cel6A cellobiohydrolases demonstrates that

the active site loops exist in two conformations, open and closed, and move up to 4.5 Å when a substrate is bound (Figure 1.3B) [32]. The movement of the active site loops is likely the basis of the endo activity, i.e. internal cleavage of cellulose, observed in Cel6A cellobiohydrolases [37].

1.4 Cellulase engineering

1.4.1 Challenges of engineering cellobiohydrolase activity

Cellulose hydrolysis is a complex reaction that occurs at the liquid-solid interface and does not obey Michaelis-Menten kinetics. The nature of two-domain cellulases is such that substrate binding is not always productive, and the cellobiohydrolase needs to diffuse along the substrate until it encounters a suitable reaction site [38]. Furthermore, the reaction rate declines rapidly during hydrolysis, leading to long hydrolysis times to achieve complete conversion or the need to increase enzyme loading to compensate for the slower rate [39]. The decline in hydrolysis rate is attributed to obstacles that are encountered on the substrate surface or other slow-moving cellulases, which bring the processivity of cellobiohydrolases to a halt [39, 40]. As a result, the reaction is limited by dissociation of the enzyme-substrate complex, where the cellobiohydrolase must unbind and locate a new substrate site before hydrolysis can continue [38]. Due to its tunnel-shaped active site, the rate of dissociation for cellobiohydrolases is low. Efforts to increase the intrinsic activity of cellobiohydrolase have made very little progress, in part due to the many parameters — substrate binding, diffusivity, processivity, and dissociation — that are carefully balanced in catalysis. In addition, cellulases are primary metabolic enzymes that are crucial to the survival of cellulolytic organisms. The cellulase activity is likely already optimized since the selection pressure for higher catalytic rates is high. Cellulolytic fungi have been reported to live at temperatures up to ∼55°C. While the optimum temperatures of the secreted cellulases can be slightly higher than the optimum temperature for growth [41], it is conceivable that thermostability has not reached its optimum peak as it has not been subjected to stringent selection pressure.

1.4.2 Advantages of thermostable cellulases

For cellulosic biomass to become a feasible feedstock for transportation fuels and chemicals, the cost of production needs to be competitive with fossil fuel production. Thermostable cellulases are desirable candidates for reducing the biomass degradation costs for several reasons. First, thermostable cellulases tend to be more stable during production, storage and over a wide range of operating conditions. Since chemical pretreatments are necessary for increasing the substrate accessibility to enzymatic degradation, thermostable cellulases are preferred as they are more resilient towards relatively harsh industrial treatments and conditions. Thermostable cellulases also allow cellulose hydrolysis to proceed at high temperature with concomitant increased cellulose degradation rate and reduced microbial contamination [42, 43]. Lastly, the process costs are lower at high hydrolysis temperature because less energy is required to bridge the reactor temperatures for pretreatment and for hydrolysis. The viscosity of hydrolysis mixture is also reduced at high temperature, lowering the energy consumption of mixing [44].

1.4.3 Previous work on Cel6A

Previously, Heinzelman et al. constructed a library of thermostable family 6 fungal cellobiohydrolases using structure-guided recombination [45]. Cel6A from *H. insolens*, *H. jecorina*, and *Chaetomium thermophilum* were divided into eight blocks and recombined. The block boundaries were identified using RASPP algorithm (Recombination as Shortest Path Problem) [46], aiming to minimize the number of disrupted side-chain contacts relative to the average number of mutations in the library. Based on the half-life data of 23 chimeras (out of 6,558 possible chimeric sequences), linear regression models were built to qualitatively evaluate the blocks' contributions to thermostability. The models classified the blocks as stabilizing, destabilizing, or neutral and identified B1P1 (block one from parent one), B6P3, B7P3, and B8P2 as stabilizing. Chimera 12222332, referred to as HJPlus in this thesis, was constructed by substituting the four stabilizing blocks into *H. jecorina* Cel6A. HJPlus Cel6A has activity close to the wild-type *H. jecorina* Cel6A but is more thermostable than the most stable parent, *H. insolens*. The high thermostability coupled with the high expression level (∼10

mg/L) in our heterologous expression host *S. cerevisiae* make HJPlus an excellent starting point for directed evolution to further increase stability.

1.5 Thesis summary

Efforts to engineer thermostable Cel6A cellobiohydrolases are detailed in this thesis. In Chapter 2, HJPlus was evolved for higher thermostability with three rounds of directed evolution, and the most thermostable variant was used to study the implication of utilizing thermostable cellulases in cellulose hydrolysis. In Chapter 3, the mechanism of thermal inactivation for the engineered thermostable Cel6A was studied by investigating the role of free cysteines in cellulase inactivation at high temperatures. In Chapter 4, thermostability measurements and ionic liquid tolerance of the Cel6A variants were examined in detail. In Chapter 5, a number of semi-rational protein designs were surveyed to further enhance the thermostability of Cel6A.

Chapter 2

Engineering Cellobiohydrolases for Higher Thermostability

2.1 Abstract

Thermostability is an important feature in industrial enzymes: it increases biocatalyst lifetime and enables reactions at higher temperatures, where faster rates and other advantages ultimately reduce the cost of biocatalysis. In this chapter, we report the thermostabilization of a chimeric fungal family 6 cellobiohydrolase (HJPlus) by directed evolution using random mutagenesis and recombination of beneficial mutations. Thermostable variant 3C6P has a half-life of 280 minutes at 75°C and a T_{50} of 80.1°C, a ~15°C increase over the thermostable Cel6A from *H. insolens* (HiCel6A) and a ~20 °C increase over that from *H. jecorina* (HjCel6A). Most of the mutations also stabilize the less-stable HjCel6A, the wild-type Cel6A closest in sequence to 3C6P. During a 60-hour Avicel hydrolysis, 3C6P released 2.4 times more cellobiose equivalents at its optimum temperature (T_{opt}) of 75°C than HiCel6A at its T_{opt} of 60°C. The total cellobiose equivalents released by HiCel6A at 60°C after 60 hours is equivalent to the total released by 3C6P at 75°C after ~6 hours, a 10-fold

reduction in hydrolysis time. A binary mixture of thermostable Cel6A and Cel7A hydrolyzes Avicel synergistically and released 1.8 times more cellobiose equivalents than the wild-type mixture, both mixtures assessed at their respective T_{opt}. Crystal structures of HJPlus and 3C6P, determined at 1.5 Å and 1.2 Å resolution, indicate that the stabilization comes from improved hydrophobic interactions and restricted loop conformations by introduced proline residues.

2.2 Introduction

High enzyme costs remain a major hurdle for the conversion of abundant and renewable cellulosic feedstocks into fermentable sugars for microbial production of fuels and chemicals [47, 48, 49]. Strategies to reduce enzyme-related costs include reducing the recalcitrance of plant substrate [50, 51], improving pretreatment conditions [15, 52], and improving enzyme performance [53, 54]. One approach to improving enzyme performance is to enhance cellulase thermostability. Cellulosic biomass hydrolysis with commercial fungal cellulase mixtures currently operates at ∼50°C; increasing hydrolysis temperature could significantly improve process performance, provided the catalysts remain active [55]. The multiple benefits of improving biocatalyst stability include extending the lifetime during production, storage, and hydrolysis and conferring higher tolerance to pretreatment chemicals [56]. More importantly, cellulose degradation at higher temperatures increases degradation rates and reduces microbial contamination [42, 43].

A fungal cellulase system for efficient cellulose degradation includes at least four glycoside hydrolases: a cellobiohydrolase to target the reducing ends of cellulose, a cellobiohydrolase to target the non-reducing ends, an endoglucanase to reduce chain length and produce free chain ends, and a β-glucosidase to convert soluble cellodextrin to glucose and relieve cellobiose inhibition for other cellulases [19, 57]. Other lytic enzymes such as the combination of copper-dependent polysaccharide monooxygenase and cellobiose dehydrogenase cleave crystalline cellulose through oxidative mechanisms and work synergistically with cellulase-mediated hydrolysis [58, 59]. Accessory proteins such as swollenins can also aid cellulose degradation by disrupting the structure of crystalline cellulose [60].

According to the CAZy (Carbohydrate-Active enZyme) database [61], the known fungal cellobio-hydrolases that act on the non-reducing ends of cellulose are found exclusively in glycoside hydrolase family 6. Despite the essential nature of Cel6A in the non-complexed fungal cellulase system, only a few protein engineering studies have reported increases in the thermostability of fungal Cel6A [45, 62, 63]. Lantz et al. (2010) at Genencor reported finding a thermostable *Hypocrea jecorina* Cel6A variant with a $\sim7°C$ increase in melting temperature (T_m) compared to the wild-type Cel6A, but the authors did not provide sequence information or describe its activity as a function of temperature. In our previous work, we used crystal structure information to construct chimeric cellulases by structure-guided SCHEMA recombination of Cel6A from *Humicola insolens*, *H. jecorina*, and *Chaetomium thermophilum* [45]. A highly thermostable Cel6A chimera, HJPlus, was constructed by incorporating three stabilizing blocks from *H. insolens* and *C. thermophilum* Cel6A into the industrially-important *H. jecorina* Cel6A. HJPlus is more thermostable than all three of its parent Cel6A enzymes and hydrolyzes more cellulose than the parents during long-time assays. In this chapter, we chose this active, thermostable chimera HJPlus as the starting point for further stabilization by directed evolution.

Wild-type cellobiohydrolases and endoglucanases from *H. jecorina* have been shown to hydrolyze cellulose synergistically as mixtures, releasing more sugar than the sum of the sugar released by the individual enzymes [64, 34]. Commercial mixture Celluclast™ derived from *H. jecorina* has been supplemented with thermostable cellulases to improve cellulose hydrolysis, but the optimum temperature of Celluclast™ mixture remains $\sim50°C$ [65, 66]. Celluclast™ has also been compared against a mixture of cellulases from different thermophilic fungi, but the novel cellulase mixture gave a similar hydrolysis yield at $60°C$ to Celluclast™ at $45°C$ [55]. To our knowledge, no study has reported fungal cellulase mixtures that hydrolyze cellulose optimally at temperatures above $65°C$. Furthermore, no one has investigated whether engineered cellulases exhibit synergy during hydrolysis at elevated temperatures.

In this chapter, we describe how the thermostability of a family 6 cellobiohydrolase was increased without compromising catalytic activity. In an effort to understand the mechanisms by which the

thermal stabilization was achieved, we solved the crystal structures of the chimera parent Cel6A and the thermostable Cel6A variant. We also show that the thermostable Cel6A hydrolyzes crystalline cellulose synergistically with a thermostable fungal Cel7A that functions optimally at 65°C [67]. A mixture of the two thermostable enzymes is more active than the wild-type mixture and hydrolyzes more crystalline cellulose during a long hydrolysis assay.

2.3   Methods

This section describes the medium-throughput screen, the cellulase assays, and the crystallization methods used in this Chapter. For materials and general methods, please refer to Chapter 6.

2.3.1   Medium-throughput cellulase screen on Avicel

2.3.1.1   Library construction

The mutant libraries were constructed by error-prone PCR, using 100 ng of plasmid, 80 nmol of dNTP, 20 pmol of forward primer, 20 pmol of reverse primer, 700 nmol of magnesium chloride, 5-20 nmol of manganese chloride, and 8 U of Taq polymerase in a 100 μL reaction. The PCR program used 30 seconds at 95°C for initial denaturation, with 20 cycles of 30 seconds at 95°C, 30 seconds at 52°C, and 72°C for 120 seconds, with 5 minutes of final extension at 72°C. The gene inserts were gel purified and extracted before transformation in S. cereviae with endonuclease digested backbone. The forward primer (5' CGG GTT ATT GTT TAT AAA TAC TAC TAT TGC CAG 3') and reverse primer (5' GAC ATG GGA GAT CGA ATT CAA CTC C 3') were 73bp upstream and 65bp downstream from the gene for the purpose of homology recombination in *S. cerevisiae*. This protocol with 100 μM manganese chloride (10 nmol of manganese chloride in 100 μL reaction) should yield ~2-3 nucleotide substitution (~1 amino acid substitution) per Cel6A gene.

Alternatively, the mutant libraries were constructed using the GeneMorph II Random Mutagenesis Kit, using 500 ng of gene insert (~2.9 ug of plasmid), 40 nmol of dNTP, 125 ng of forward and of reverse primer, and 2.5 U of Mutazyme II polymerase in a 50 μL reaction. The PCR program used

2 min at 95°C for initial denaturation, and 25 cycles of 30 seconds at 95°C, 30 seconds at 50°C, and 72°C for 120 seconds, with 10 minutes of final extension at 72°C. For each library construction, four 50 μL PCR reactions were set up in parallel. After PCR, the 200 μL PCR reactions were digested with 2 μL of DpnI at 37°C for 2 hours before the gene inserts were gel purified and extracted. This protocol with the amount of gene insert and cycle number specified should yield $\sim$3 nucleotide substitution ($\sim$2 amino acid substitution) per Cel6A gene.

The recombination library was constructed via overlap extension PCR as described previously [68].

2.3.1.2 Cellulase expression in *S. cerevisiae* in 96-well plates

To express the mutagenesis libraries in *S. cerevisiae* strain YDR483W BY4742 (ATCC No. 4014317), a high-efficiency transformation protocol adapted from Chao et al. [69] was used. A single colony was used to inoculate 5 mL of YPD medium and grown overnight at 30°C and 250 rpm in an orbital shaker. The Abs_{600} of the overnight culture was measured (typically with 1:50 dilution in water), and the overnight culture was used to inoculate 20 mL of YPD medium to an Abs_{600} of 0.1. The YPD culture was grown at 30°C for $\sim$6 hours until the Abs_{600} reached $1.3 - 1.5$ (typically measured with 1:5 dilution in water). Once the cells reached the desired density, the culture was mixed with 200 μL of 1 M Tris buffer, pH 8.0, with 2.5 M DTT and shaken at 30°C and 250 rpm for 15 minutes. Cells were harvested by centrifugation at 2,500 x g at 4°C for 3 minutes, resuspended with 10 mL ice-cold buffer E (10 mM Tris buffer, pH 7.5, with 0.27 M sucrose and 2 mM magnisum chloride), centrifuged again to be resuspended with 1 mL of ice-cold buffer E, and centrifuged a third time to be resuspended with 200 μL of buffer E. Samples containing 50 μL of cell solution and 5 μL of 0.5 μg backbone mixed with 0.5 μg of insert were electroporated at 0.54 kV, 25 μF, and infinite resistance in 0.2 cm electroporation cuvettes (Biorad). The electroporated samples were immediately rescued with 1 mL of warm YPD and recovered at 30°C and 250 rpm for 1 hour. The transformed cells were plated based on an average transformation efficiency of 10 colonies per μL of recovered cells and grown for 3 days at 30°C. Colonies containing mutant Cel6A were used to inoculate 50 μL SD-Ura

medium in 96-well plates covered with AirPore sheet (Qiagen). The cultures were grown overnight at 30°C and 80% humidity in orbital shakers before the addition of 350 μL YPD medium and growth for an additional 48 hours. The culture was harvested by centrifugation, and the supernatant was used for activity assays without further treatment.

2.3.1.3 Medium-throughput Cel6A activity assay

Three-day culture supernatants (100 μL) were combined with 3 mg of Avicel in 96-well PCR plates and incubated at 4°C for 1.5 hours. Avicel and the bound enzymes were washed three times with 180 μL of 50 mM sodium acetate buffer, pH 5.0, before resuspended in 75 μL of buffer. The reaction was incubated in 75°C waterbath for two hours and then cooled on ice. After hydrolysis, 50 μL of reaction supernatants were sampled for reducing sugar concentrations via Nelson-Somogyi sugar assay. The library hits were streaked on SD-Ura agar plates and grown at 30°C for 3 days. Single colonies of the library hits were used to inoculate a new 96-well plate of culture and rescreened for activity to verify performance over parental enzyme. To recover plasmid from library hits, $10 - 50$ μL of resuspended cells were lysed and recovered using Zymoprep Yeast Plasmid Miniprep II Kit.

2.3.2 Thermostability measurements

2.3.2.1 Half-life measurements

All half-life measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. Samples containing 2 μg of Cel6A in 40 μL were inactivated at 75°C for up to 5 hours in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. After heat inactivation, the enzymes were incubated at 50°C for 2 hours with 60 μL of 5 % w/v Avicel to measure the residual activity. Half-lives were determined from plots of the natural logs of residual activities versus the inactivation time. Reported values were averaged from at least three independent measurements.

2.3.2.2   T$_{50}$ measurements

All T$_{50}$ measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. Samples containing 2 μg of Cel6A in 40 μL were inactivated at different temperatures for 15 minutes in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. The temperature range was selected to ensure the T$_{50}$ value is bracketed by a 20°C range. After heat inactivation, the enzymes were incubated at 50°C for 2 hours with 60 μL of 5% w/v Avicel to measure the enzyme's residual activity. To determine T$_{50}$, the residual activities were plotted against the temperature using SigmaPlot (Systat Software Inc) and fitted using the 4-parameter Botlzmann sigmoidal function ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$). T$_{50}$ is the inactivation temperature with half maximal residual activity, or x$_0$. Reported values were averaged from at least three independent measurements.

2.3.3   Cellulase activity measurements

All cellulase activity measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. To determine activity-temperature profiles of Cel6A, samples containing 2 μg of purified Cel6A were combined with 3 mg of Avicel in 100 μL and incubated at 40°C to 90°C for 2 hours or at 60°C to 75°C for 60 hours in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. To determine the activity of the Cel6A and Cel7A mixture, purified Cel6A and Cel7A were combined at different ratios to a final concentration of 0.5 μM along with 3 mg of Avicel in 100 μL and incubated at 50°C to 70°C for 60 hours. After hydrolysis, the reaction supernatants were sampled for reducing sugar concentrations via Nelson-Somogyi assay using cellobiose as the reducing sugar standard.

2.3.4   Cellulose adsorption measurements

Samples containing 2 μg to 8 μg of Cel6A in 200 μL of 50 mM sodium acetate buffer, pH 5.0, were incubated with or without 3% w/v Avicel at 4°C for 90 minutes. The protein concentrations of the supernatants were determined using the Bradford assay. The measurements with and without Avicel at different initial loading concentrations were used to determine the fraction of protein adsorption to Avicel.

2.3.5 Crystallography

2.3.5.1 Crystallization and X-ray data collection

For crystallization, the catalytic domains of HJPlus (HJPlus-cat) and 3C6P (3C6P-cat) with C-terminal His6-tag were cloned and expressed in yeast. The catalytic domains were purified using HisTrap HP columns and then HiTrap Q HP columns (GE Healthcare). The enzymes were deglycosylated with PNGase F as described in Chapter 6. HJPlus-cat was crystallized at 25°C by the sitting drop vapor diffusion method, using 34% poly(ethylene glycol) 1000, 200 mM zinc acetate, and 100 mM sodium acetate, pH 5.75. 3C6P-cat was also crystallized at 25°C by the sitting drop vapor diffusion method, using 38% poly(ethylene glycol) 1000, 100 mM ammonium sulfate, and 100 mM sodium acetate, pH 6.0. A cryoprotectant solution was made by including 25% v/v glycerol in the mother liquor and added to the crystals before they were looped and flash frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K at the Stanford Synchrotron Radiation Lightsource, beamline 12-2 on a Dectris Pilatus 6M detector. Diffraction datasets were integrated with XDS [70] and scaled using SCALA [71].

2.3.5.2 Structure determination and refinement

For HJPlus-cat, initial phases were determined using molecular replacement against the wild type *H. insolens* structure 1OCN, chain A [37]. The initial phases of 3C6P-cat were determined using molecular replacement against the structure model of HJPlus-cat, while maintaining the Rfree statistics of HJPlus-cat to prevent over-fitting the data. Molecular replacement was accomplished using MOL-REP [72] within the CCP4 software suite [73]. Refinement was accomplished with iterative cycles of manual model building within COOT [74] and automated refinement using REFMAC [75] within CCP4. Final cycles of REFMAC refinement included TLS parameters. All protein structure figures were generated using PyMol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.).

2.4 Results

2.4.1 Directed evolution for increased thermostability

Beginning with the thermostable fungal Cel6A chimera HJPlus [45], we identified further stabilizing mutations by random mutagenesis and screening for improved total activity on crystalline cellulose (Avicel) at 75°C over a 2-hour period. The mutant cellobiohydrolase libraries were expressed and secreted by *S. cerevisiae* in 96-well plate format, and Cel6A activity was quantified using the Nelson-Somogyi reducing sugar assay. Because the reducing sugar assay does not distinguish between the sugar released during hydrolysis and the sugar present in the culture medium, we eliminated the background from the medium by first batch-purifying the mutated enzymes on Avicel, utilizing the fact that the cellulose binding module (CBM) from *H. jecorina* Cel6A binds to Avicel irreversibly at 4°C [76], before proceeding with the activity assay. Approximately 2,800 colonies from the HJPlus random mutagenesis library were screened; improved variants exhibited up to 40% higher activity than the parent. The best-performing variant 1G6 had a single amino acid substitution, S317P. The second-generation random mutagenesis library was constructed using 1G6 as the template. Screening identified variants with up to 60% higher activity than the parent 1G6. The best-performing variant 2B3 incorporated an additional mutation, leading to Q277L. No synonymous mutations were found.

Other mutations identified in the top five variants encoded substitutions S30F, V128A, V131E, S293R, and S413F from the first generation, and M135L, S406P, S413P, and S413F from the second generation. A third-generation library using variant 2B3 as the template was therefore constructed to recombine potentially beneficial mutations at positions S30, V128, V131, M135, S293, S406, and S413. Two amino acids were allowed at each position (three in the case of S413), wild-type or the beneficial mutation discovered in the first- and second-generation libraries. The best variant 3C6P found from this library after three-fold oversampling contained a total of seven mutations from HJPlus: S30F, V128A, M135L, Q277L, S317P, S406P, and S413P.

2.4.2   Characterization of thermostabilized HJPlus variants

The chimera HJPlus and its improved first-, second-, and third-generation variants 1G6, 2B3, and 3C6P were purified to homogeneity using standard Ni-NTA columns, as were the wild-type Cel6A parent enzymes from *H. insolens*, *H. jecorina*, and *C. thermophilum*. The "wild-type" Cel6A enzymes in this report are all recombinant enzymes expressed in *S. cerevisiae* as described in the Methods section. Multiple sequence alignments of the wild-type Cel6 enzymes and 3C6P can be found in Appendix A. Table 2.1 compares the thermostabilities of all these Cel6A enzymes using two different metrics, the half-life at 75°C and the T_{50}. The half-life at 75°C assesses how long an enzyme remains active when incubated at this temperature, but is not very informative when the enzyme inactivates at a lower temperature. The T_{50} measurement estimates the temperature at which 50% of the enzyme is inactivated after 15 minutes of heat incubation and allows direct comparisons of enzymes with very different thermostabilities. Both *H. insolens* (HiCel6A) and *H. jecorina* Cel6A (HjCel6A) were completely inactivated within five minutes at 75°C, while the half-life of *C. thermophilum* Cel6A (CtCel6A) could not be determined because the enzyme unfolds at 75°C but refolds (partially) when the temperature is reduced. The three wild-type Cel6A have T_{50} values between 60.0°C and 65.2°C. Thermostable chimera HJPlus unfolds irreversibly and has a half-life of 8.8 minutes at 75°C and a T_{50} at 71.9°C.

Directed evolution of HJPlus to increase total activity at an elevated temperature led to a significant increase in thermostability. The half-lives of the Cel6A variants at 75°C increase more than 30-fold, from 8.8 minutes for HJPlus to 280 minutes for 3C6P, while the T_{50} increases by 8.2°C, from 71.9°C to 80.1°C (Table 2.1). All of these Cel6A variants undergo irreversible unfolding and exhibit first-order thermal inactivation kinetics. Example residual activity plots for determining the half-life and T_{50} measurements of HJPlus, 3C6P, and wild-type Cel6A can be found in Appendix B.

The thermostability screen involves partial purification by binding to Avicel at low temperature. Mutants that bind more tightly could therefore appear to have higher activity in the screen. We thus tested adsorption to Avicel for HJPlus and variant HJPlus S30F at 4°C with enzyme concentrations similar to those used to measure cellobiohydrolase activity. As shown in Figure 2.1, the fraction of

Table 2.1: Thermostabilities of wild-type and engineered Cel6A. HiCel6A, HjCel6A, and CtCel6A are wild-type Cel6A enzymes from *H. insolens*, *H. jecorina*, and *C. thermophilum*, while HJPlus, 1G6, 2B3, and 3C6P are engineered Cel6A.

Enzyme	Mutation(s) wrt HJPlus	$t_{1/2,75°C}$ (min)	T_{50} (°C)
HiCel6A	97 mutations	≤ 2.5	65.2 ± 0.1
HjCel6A	48 mutations	≤ 2.5	60.2 ± 0.4
CtCel6A	84 mutations	N.D.	62.6 ± 0.4
HJPlus	-	8.8 ± 0.6	71.9 ± 0.6
1G6	S317P	14 ± 3	73.2 ± 0.3
2B3	Q277L+S317P	37 ± 4	75.7 ± 0.3
3C6P	S30F+V128A+M135L	280 ± 13	80.1 ± 0.4
	+Q277L+S317P+S406P+S413P		

HJPlus S30F bound to Avicel is higher than that of HJPlus, indicating that HJPlus S30F has a higher affinity for Avicel than HJPlus.

Lastly, we investigated whether the mutations affected cellobiohydrolase activity at 50°C, a temperature at which HJPlus and its variants do not denature significantly during the hydrolysis assay. At 50°C, HJPlus and its variants all have activities that are comparable to the wild-type Cel6A enzymes (Figure 2.2).
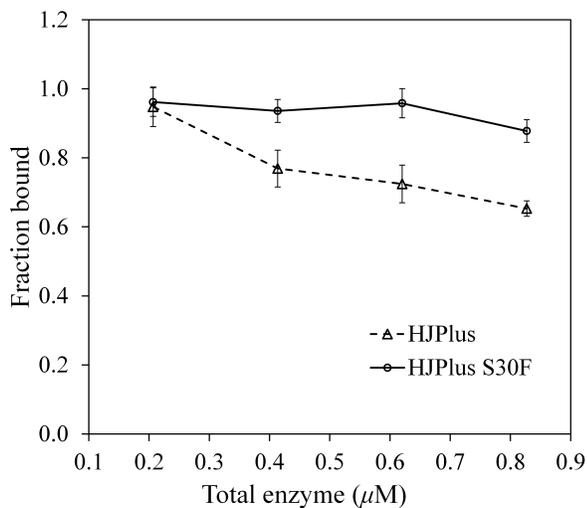
Figure 2.1: Adsorption of HJPlus and HJPlus S30F at 4°C with 3% w/v Avicel.
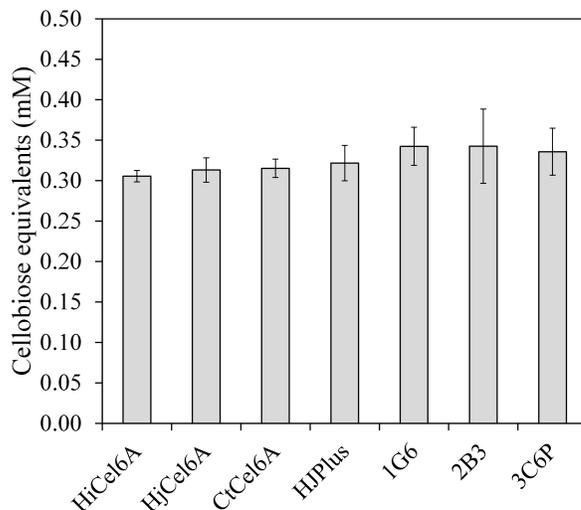
Figure 2.2: Activities of wild-type and engineered Cel6A at 50°C. HiCel6A, HjCel6A, and CtCel6A are wild-type Cel6A from *H. insolens*, *H. jecorina*, and *C. thermophilum*. Purified Cel6A were assayed for 2 hours with 3% w/v Avicel in 50 mM sodium acetate buffer, pH 5.0. Activities are reported as the total cellobiose equivalents released.

2.4.3 Stabilizing mutations in *H. jecorina* Cel6A

We sought to determine whether these stabilizing mutations are also stabilizing in HjCel6A. With the exception of Q276L (equivalent to Q277L in 3C6P; see Table 2.2), all the mutations in the catalytic domain (M134L, S316P, S406P, S413P) are stabilizing in HjCel6A, with an increase of T_{50} up to 3.6°C (Table 2.2). S30F does not change the T_{50} of HjCel6A, but its higher affinity for Avicel is expected to translate to HjCel6A, which shares the same CBM and linker with the HJPlus variants.

2.4.4 Temperature-activity profiles of Cel6A

Directed evolution has increased the thermostability of HJPlus, as measured by half-life and T_{50}, at no cost to the enzyme activity at 50°C. We next determined how this enhanced thermostability translates to hydrolysis performance at high temperatures. First we examined the cellobiohydrolase activities of HJPlus and 3C6P after 2-hour Avicel hydrolysis from 40°C to 90°C and compared those with the activities of HiCel6A, HjCel6A, and CtCel6A, which are some of the closest wild-type Cel6A in sequence to HJPlus. With an optimum temperature (T_{opt}) between 60°C and 65°C, HiCel6A

has the highest T_{opt} and the highest total activity at its T_{opt} of the three wild-type Cel6A enzymes tested (Figure 2.3A). For this reason, all further comparisons with the engineered enzymes were made to HiCel6A. Chimera HJPlus (which differs from HiCel6A at 97 positions) has a T_{opt} of 70°C, an increase of roughly 5°C. The 7 additional mutations in 3C6P increase T_{opt} by another 5°C, to 75°C, a temperature at which all the wild-type Cel6A enzymes have only minimal activities. In 2 hours, 3C6P hydrolyzed 50% more sugar than HiCel6A and 20% more than HJPlus acting at their respective T_{opt}.

The advantages of greater stability become even more significant during longer hydrolysis experiments, where the thermostable enzyme is both more active and has a longer lifetime. To demonstrate this, we tracked Avicel hydrolysis by HiCel6A and 3C6P over 60 hours, at their respective T_{opt}, 60°C and 75°C. As shown in Figure 2.3B, 3C6P hydrolyzed 1.6 times more sugar at 60°C than HiCel6A after 60 hours, which can be attributed to the longer half-life of 3C6P at this temperature. At 75°C, HiCel6A had minimal activity, while 3C6P retained activity throughout the 60-hour period and hydrolyzed 2.4 times more sugar than HiCel6A did at 60°C. In addition, the total cellobiose equivalents released by HiCel6A at 60°C after 60 hours is equivalent to that by 3C6P at 75°C after approximately 6 hours, a 10-fold reduction in hydrolysis time, or by 3C6P at 60°C after approximately 19 hours, a more than 3-fold reduction. In both comparisons, using a more thermostable Cel6A enzyme can reduce loading or hydrolysis time, both of which profoundly affect enzyme costs.

Table 2.2: T_{50} of *H. jecorina* Cel6A (HjCel6A) variants containing mutations found in 3C6P.

Enzyme variant	Equivalent mutation in 3C6P	T_{50} (°C)
HjCel6A	-	60.2 ± 0.4
HjCel6A S30F	S30F	60.3 ± 0.3
HjCel6A M134L	M135L	62.0 ± 0.5
HjCel6A Q276L	Q277L	58.6 ± 0.4
HjCel6A S316P	S317P	61.9 ± 0.2
HjCel6A S406P	S406P	61.5 ± 0.3
HjCel6A S413P	S413P	63.8 ± 0.4
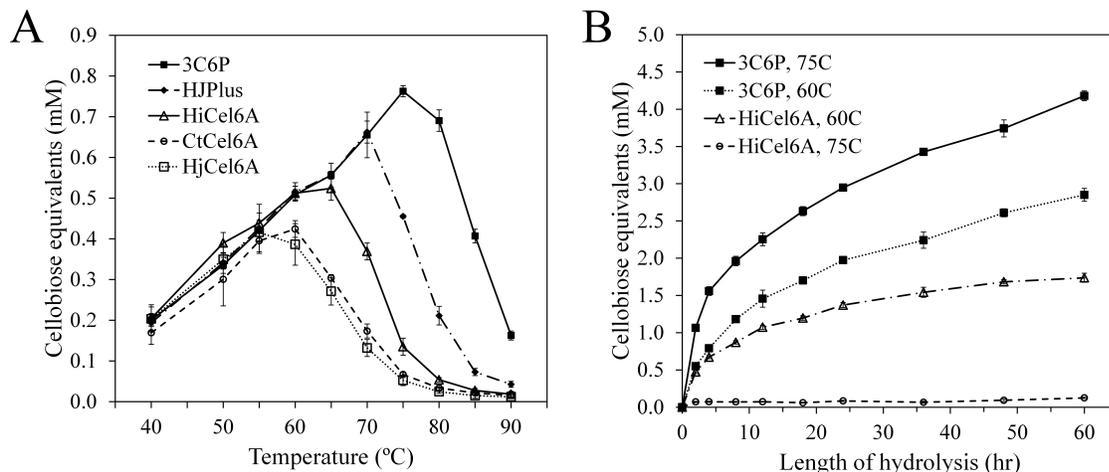
Figure 2.3: Activity-temperature profiles of Cel6A enzymes. HiCel6A, CtCel6A, and HjCel6A are wild-type Cel6A from *H. insolens*, *C. thermophilum*, and *H. jecorina*, while 3C6P and HJPlus are engineered Cel6A variants. A) Total cellobiose equivalents released at 40°C to 90°C after 2-hour Avicel hydrolysis. B) Total cellobiose equivalents released at 60°C and 75°C during 60-hour Avicel hydrolysis. Purified Cel6A were assayed for 2 hours or 60 hours with 3% w/v Avicel in 50 mM sodium acetate buffer, pH 5.0. Activities are reported as the total cellobiose equivalents released.

2.4.5 Exo-exo synergy between Cel6A and Cel7A

Fungal Cel7A is a cellobiohydrolase that acts on the cellulose reducing end. In our previous work, we engineered Cel7A using recombination [77] and predictive methods [67] to construct a thermostable Cel7A variant TS8 that is optimally active at 65°C, with a T_{50} that is 9.2°C higher than the thermostable wild-type Cel7A from *Talaromyces emersonii* (TeCel7A). We examined whether wild-type and engineered-thermostable Cel6A and Cel7A enzymes hydrolyze cellulose synergistically, the foundation of efficient cellulose degradation. The total cellobiose equivalents released by thermostable wild-type Cel6A and Cel7A (HiCel6A and TeCel7A) and the engineered-thermostable Cel6A and Cel7A (3C6P and TS8), both as mixtures and as individual enzymes, were assessed after 60 hours at 50°C, 60°C, and 70°C (Figure 2.4). The wild-type and engineered-thermostable Cel6A and Cel7A mixtures both hydrolyze cellulose synergistically, releasing up to 2-fold more cellobiose equivalents when acting as a mixture compared to the sum of the cellobiose equivalents released by the individual enzymes. The mixture of 3C6P and TS8 is more active than the mixture of HiCel6A and TeCel7A at both 60°C and 70°C, with the engineered thermostable mixture releasing 1.8 fold more
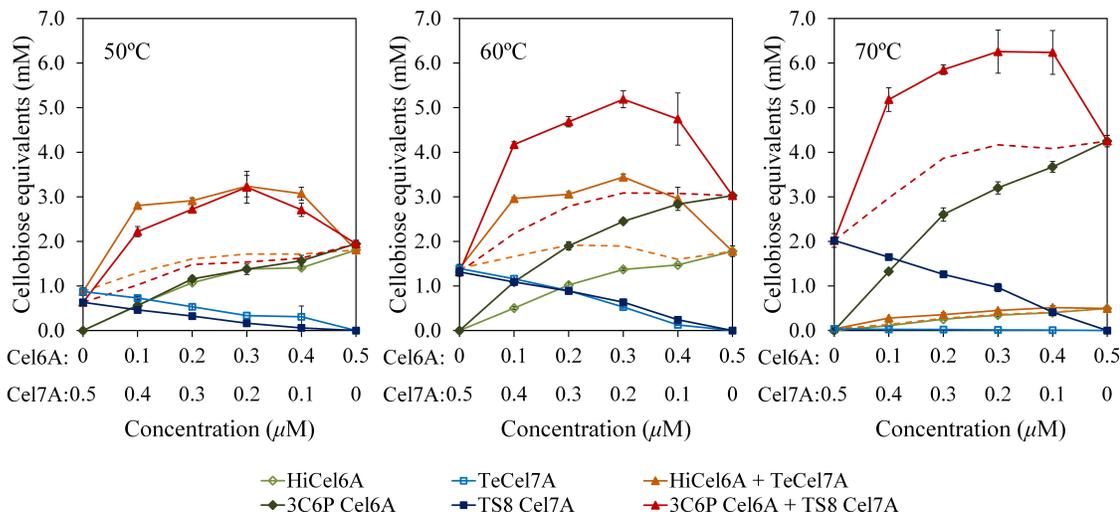
Figure 2.4: Synergistic cellulose hydrolysis by wild-type and engineered-thermostable Cel6A and Cel7A. HiCel6A is wild-type Cel6A from *H. insolens*; TeCel7A is wild-type Cel7A from *T. emersonii*; 3C6P is an engineered thermostable Cel6A, and TS8 is an engineered thermostable Cel7A. Total cellobiose equivalents released from 0.5 μM of purified Cel6A and Cel7A combined at different molar ratios were assessed after 60 hours at 50°C, 60°C, and 70°C. The cellobiose equivalents released by Cel6A or Cel7A alone are included as controls. The dotted lines are the sums of cellobiose equivalents release from the individual enzymes.

sugar at its T_{opt} of 70°C than the wild-type mixture at its T_{opt} of 60°C. The preferred molar ratio in the thermostable mixture also trends towards higher Cel6A concentration as the temperature increases, favoring the mixture with higher concentration of 3C6P, which is more thermostable than TS8 Cel7A.

2.4.6 Crystal structures of HJPlus and 3C6P

To gain insight into the molecular basis of the significant increase in thermostability conferred by the beneficial mutations, we solved the X-ray crystal structures of HJPlus and 3C6P to a resolution of 1.5 Å (PDB ID 4I5R) and 1.2 Å (PDB ID 4I5U), respectively. Statistics for data collection and final protein structure models are given in Table 2.3. The structures of HJPlus and 3C6P both display a distorted $(\beta/\alpha)_7$ barrel fold typical of family 6 glycoside hydrolases. HJPlus shares 87% sequence identity with HjCel6A in the catalytic domain; superposition of the HJPlus and HjCel6A structure 1QK2, chain A, reveals a low backbone root mean square deviation (rmsd) of 0.32 Å, indicating that

the chimeragenesis did not significantly alter the overall structure of the enzyme. HJPlus and its variant 3C6P have a backbone rmsd of only 0.24 Å. In both cases, the mutations cause only subtle changes in the structures.

The five thermostabilizing mutations in 3C6P are all near the surface of the enzyme, but only the serine-to-proline mutations are solvent-exposed. Mutation M135L resides near the end of the substrate tunnel and is surrounded by four aromatic residues as well as other hydrophobic residues (Figure 2.5). In addition, surface mapping in the structure of HJPlus and 3C6P by a probe radius of 1.4 Å demonstrates that there are cavities near M135 in HJPlus, which disappear after the

Table 2.3: Data collection and refinement statistics for HJPlus and 3C6P crystals. All data sets were collected from single crystals. Highest-resolution shell is shown in parentheses.

	HJPlus	3C6P
PDB ID	4I5R	4I5U
Data collection		
Space group	C 1 2 1	C 1 2 1
Wavelength	1.033	1.033
Cell dimensions		
a, b, c (Å)	157.61, 45.26, 58.60	158.60, 45.41, 58.19
β (°)	95.21	95.47
Resolution (Å)	35.3 - 1.5 (1.5 - 1.6)	35.3 - 1.22 (1.22 - 1.29)
Rmerge (%)	4.1(36.5)	2.7(21.5)
I / σI	16.1(3.1)	16.3(3.7)
Completeness (%)	97.9(97.7)	87.3(71.2)
Redundancy	3.4(3.2)	2.1(1.9)
Refinement		
Resolution (Å)	35.3 - 1.5	35.3 - 1.22
No. reflections	61842	101502
R_{work} / R_{free}	0.15/0.18	0.14/0.16
No. atoms		
Protein	2795	2833
Ligand/ion	47	76
Water	278	327
B-factors		
Protein	16.3	9.78
Ligand/ion	36.4	30.5
Water	28.4	23.4
R.M.S. deviations		
Bond lengths (Å)	0.028	0.030
Bond angles (°)	2.52	2.61
Ramachandran map		
In preferred regions (%)	95.3	95.8
In allowed regions (%)	4.7	3.9
Outliers (%)	0.0	0.3
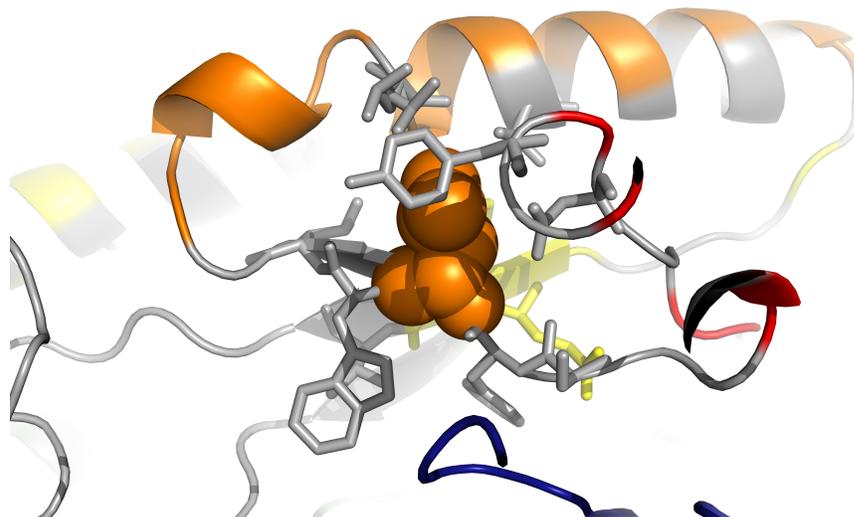
Figure 2.5: Local environment near M135L in 3C6P structure. Mutation M135L is surrounded by conserved residues among the three parents and residues from *H. jecorina*. M135L is shown as balls, and the residues within a 4 Å radius from the mutation are shown as sticks. The conserved residues are colored grey, while the non-conserved residues are colored according to the recombination blocks. Red represents block 1 from *H. insolens*. Orange and yellow represent block 2 and 3 from *H. jecorina*. Indigo represents block 7 from *C. thermophilum*. Block definitions are from the work by Heinzelman et al. [45] and can be found in Appendix C.
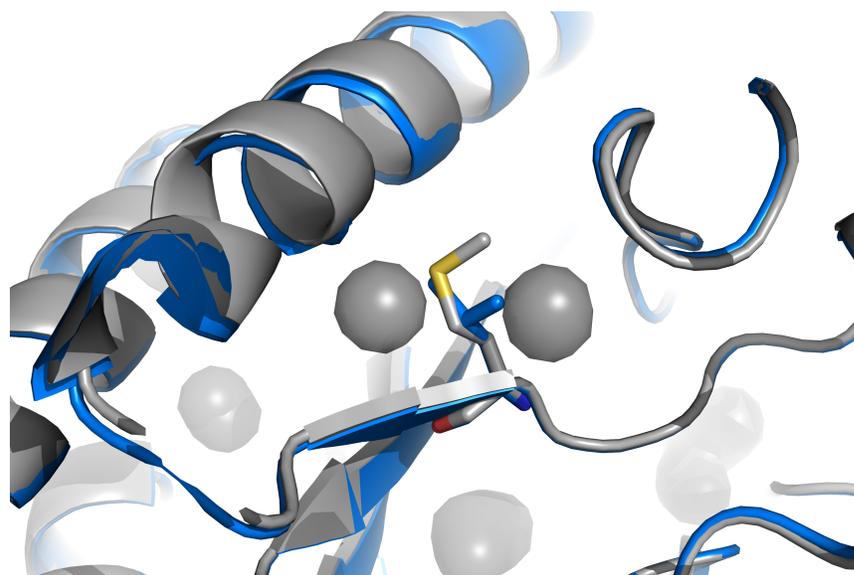
Figure 2.6: Surface mapping of the HJPlus and 3C6P structures. Residues at position 135 are shown in sticks in HJPlus (grey) and 3C6P (blue). The two grey spheres shown in the center represent cavities that are present only in HJPlus.

leucine substitution (Figure 2.6). The increase in thermostability is likely due to an increase in hydrophobicity and improved side chain packing. Mutation M135L is in a sequence block taken from *H. jecorina* (the block definitions are based on the work by Heinzelman et al. [45] and can be found in Appendix C) and is surrounded by residues that are conserved among the three parents (Figure 2.6). None of the parents have L at this position. The non-conserved residues within 4 Å of M135L are all wild-type residues from *H. jecorina*. Since M135L is surrounded by the same residues that surround M134 from *H. jecorina*, the thermostabilization effect of M135L in 3C6P translates well in HjCel6A .

In contrast, mutation Q277L is found to be stabilizing in 3C6P but not in HjCel6A. A close examination of the Q277L local environment in 3C6P reveals that Q277L resides in an α-helix in block 5 from *H. jecorina*, and it is adjacent to an α-helix in block 6 from *C. thermophilum* (Figure 2.7). Q277 is not conserved among all three parents and is I279 in *C. thermophilum*. Q277 is surrounded by conserved residues within a 4 Å radius, with the exception of F334 in block 6 from *C. thermophilum*, which is an isoleucine (I333) in *H. jecorina*. The interaction between Q277 and F334 in HJPlus is a novel contact arising from recombination, and the mutation Q277L in 3C6P restores the hydrophobic side chain interaction between (iso)leucine and phenylalanine that is observed in the *C. thermophilum* enzyme. Q276L in HjCel6A is not stabilizing because the same (iso)leucine-phenylalanine interaction does not exist at the corresponding residues in HjCel6A.

Three thermostabilizing mutations — S317P, S406P, and S413P — are serine to proline mutations in loop regions. Close examination of the mutations indicates that the substitutions of serine with proline maintain the local loop structures (Figure 2.8). Strikingly, the Cα and Cβ atoms of the serines in HjCel6A and HJPlus align remarkably well with the Cα and Cβ atoms from the corresponding prolines in 3C6P. The thermostability enhancement is likely due to the limited conformational freedom the prolines provide without straining the backbone structure.
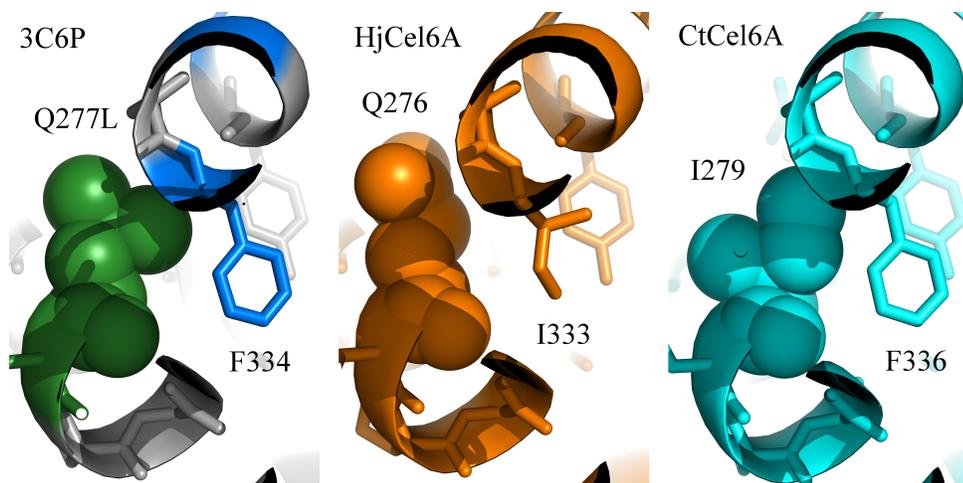
Figure 2.7: Local environments near Q227L in 3C6P structure and the corresponding residues in HjCel6A structure and CtCel6A structure. Mutation Q277L and the corresponding residue in HjCel6A and CtCel6A are shown as balls, while residues within a 4 Å radius from the mutation are shown as sticks. The side chain interaction between Q277L and F334 are shown in 3C6P, and the corresponding residues in HjCel6A structure 1QK2 (orange) and CtCel6A structure 4A05 (cyan) are displayed in parallel. In 3C6P, the conserved residues are colored grey, and the non-conserved residues are colored according to the recombination blocks. Green represents block 5 from *H. jecorina*, and blue represents block 6 from *C. thermophilum*. Block definitions are from the work by Heinzelman et al. [45] and can be found in Appendix C.

2.5   Discussion

Fungal cellulases are attractive candidates for optimization because industrially relevant fungal expression hosts such as *H. jecorina* are known to express cellulases at high levels — 100 grams per liter and possibly more [53, 78]. Optimal cellulose hydrolysis can only be achieved through synergistic actions of cellobiohydrolases, endoglucanases, and β-glucosidases, as well as other lytic enzymes, and improving the activity and thermostability of the individual components remains much desired. Several wild-type and engineered cellulases have been reported to be active above 70°C. Parry et al. reported wild-type Cel5 purified from *T. aurantiacus* to be optimally active at 70 – 80°C [79]. Heinzelman et al. [77], Komor et al. [67], and Dana et al. [80] have all reported engineered thermostable Cel7A homologous to *T. emersonii* Cel7A that are active at 70°C, though their optimum temperatures are 60 – 65°C. In the present work, we have further enhanced the thermostability of Cel6A to a T_{opt} of 75°C and demonstrated the advantage of the thermostable variant over its wild-type counterparts. The thermostable variant can achieve the same overall cellobiose equiv-
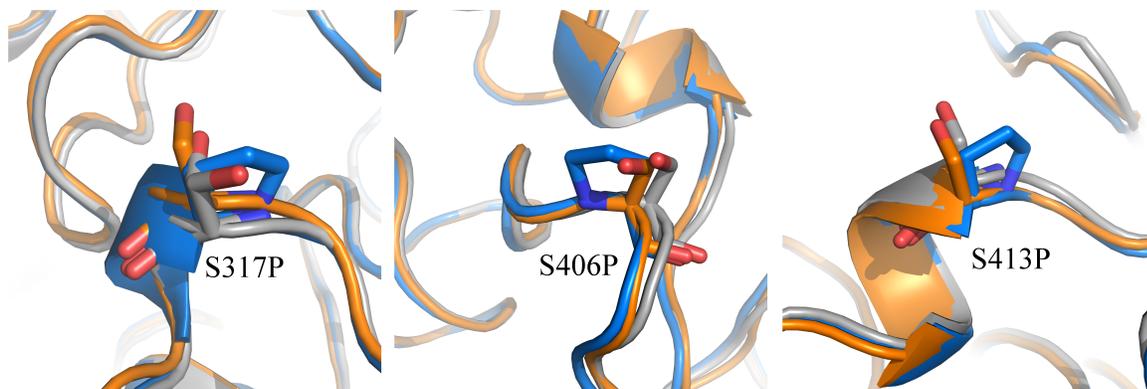
Figure 2.8: Local environments near the serine-to-proline substitutions in HjCel6A structure 1QK2 (orange), HJPlus (grey), and 3C6P (blue).

alent production as the wild-type enzymes at a small fraction of the hydrolysis time and release significantly (up to 2.4 fold) more cellobiose equivalents during long-time hydrolysis reactions.

However, improvement of a single cellulase is irrelevant if performance of the cellulase mixture is not improved. Wild-type fungal cellulase mixtures hydrolyze cellulose synergistically, releasing more sugar than the sum of individual activities [19]. The cooperative activity of the cellulase mixture is based on endo-exo synergy, exo-exo synergy, and the reduction of cellobiose inhibition by β-glucosidase. The basis of exo-exo synergy, however, remains elusive [81, 40]. The degree of synergistic effect decreases with longer hydrolysis time [34] and with saturating enzyme loadings [64]. Here we have shown that the exo-exo synergy holds for cellobiohydrolases from different fungi. The degree of synergistic effect found in this study, $\sim$2 fold, is higher than that of the *H. jecorina* Cel6A-Cel7A synergy on Avicel ($\sim$1.2 fold after 48 hours) [34] and that of the *H. insolens* Cel6A-Cel7A synergy on bacterial cellulose ($\sim$1.5 fold after 24 hours) [22].

We obtained the crystal structures of HJPlus and 3C6P catalytic domains to infer the mutations' stabilizing mechanisms. The structure models indicate a high level of structural conservation with *H. jecorina* Cel6A, despite a difference in T_{50} of $\sim$20°C. High structural similarity among enzymes of very different thermostabilities has been observed in both laboratory-evolved [82] and wild-type enzymes [83]. Despite numerous studies applying observed stabilizing factors from hyperthermophilic enzymes, the apparent lack of correlation between thermostability and structural motifs demonstrates the difficulty in pinpointing a universal recipe for stabilizing enzymes.

In conclusion, we engineered a highly thermostable Cel6A variant and demonstrated that enhancing the thermostability of cellulases is an effective strategy for improving the efficiency of cellulose hydrolysis. Engineered thermostable Cel6A and Cel7A retain synergies similar to wild-type mixtures, illustrating the utility of thermostabilizing individual cellulases to produce novel thermostable mixtures. The ability to combine the most active or thermostable cellulases regardless of the origin of the enzymes supports the approach of engineering individual components separately. The modular nature also opens up the possibility of designing novel cellulase mixtures for different hydrolysis requirements.

Chapter 3

Role of Cysteine Residues in Thermal Inactivation

3.1 Abstract

Numerous protein engineering studies have focused on increasing the thermostability of fungal cellulases to improve production of fuels and chemicals from lignocellulosic feedstocks. However, the engineered enzymes still undergo thermal inactivation at temperatures well below the inactivation temperatures of hyperthermophilic cellulases. In this report, we investigated the role of free cysteines in the thermal inactivation of wild-type and engineered fungal family 6 cellobiohydrolases (Cel6A). The mechanism of thermal inactivation of Cel6A is consistent with disulfide bond degradation and thiol-disulfide exchange. Circular dichroism spectroscopy revealed that a thermostable variant lacking free cysteines refolds to a native-like structure and retains activity after heat treatment over the pH range 5 – 9. Whereas conserved disulfide bonds are essential for retaining activity after heat treatment, free cysteines contribute to irreversible thermal inactivation in engineered thermostable Cel6A as well as Cel6A from *H. jecorina* and *H. insolens*.

3.2 Introduction

Traditionally employed in the textile, detergent, or pulp refining industries, cellulases have recently received attention for applications in production of biofuels from lignocellulosic feedstocks [19, 84, 85].

Fungal cellulases from *Hypocrea jecorina* are attractive because of their high expression titers of up to 100 grams per liter [78]. Various approaches to improving the thermostability of cellulases for industrial applications have been published [82, 45, 86, 80, 67, 87, 88]. Nonetheless, thermostabilities of engineered fungal cellulases still fall short when compared to hyperthermophilic cellulases such as those isolated from the anaerobic bacterium *Thermotoga maritima* [89]. These hyperthermophilic cellulases, however, tend to have low activity on crystalline cellulose [90] and lack high expression yields. Little effort has been devoted to understanding the mechanisms by which fungal cellulases become thermally inactivated, even though such information might assist in the generation of hyperthermophilic versions.

Mechanisms of irreversible protein thermal inactivation include deamidation of asparagine and glutamine residues, peptide chain cleavage, and cysteine oxidation [89, 91]. Cysteine is one of the least abundant amino acids in proteins but is also one of the most versatile, coordinating metals for catalysis and participating in redox regulation [92]. Whereas strategically-placed cysteines can be oxidized to form disulfide bonds, which impose structural rigidity crucial to protein stability and activity, free thiols are susceptible to chemical modification and thiol-disulfide exchange, which can lead to misfolding and aggregation [93, 94]. Free cysteines are similarly complex in their effects, as both removal and strategic placement of cysteine have been shown to enhance thermostability [95, 62, 96]. Given the reactivity of cysteine, optimizing the content of disulfide bonds and free cysteines in an enzyme can have profound implications for industrial applications [97, 93].

Cellobiohydrolase II from *H. jecorina* belongs to glycoside hydrolase family 6 (Cel6A) and is an essential enzyme for the degradation of cellulose for biofuel production [98, 34, 35]. The crystal structure of *H. jecorina* Cel6A reveals a distorted $(\beta/\alpha)_7$ barrel fold with a tunnel-shaped active site [21, 36]. The active site is composed primarily of surface loops, and the N-terminal loops and the C-terminal loops each contain one of the two conserved disulfide bonds (Cys176-Cys235 and Cys368-Cys415 in *H. jecorina* Cel6A) [99, 100]. Whereas the disulfide bonds are presumed to provide loop stabilization important for substrate binding [101, 99], the role of free cysteines in Cel6A is less clear. Previously, we engineered Cel6A cellobiohydrolases related to *H. jecorina* Cel6A using

structure-guided recombination [45] and random mutagenesis [88] to improve thermostability. The most thermostable Cel6A variant obtained, 3C6P, has five cysteine residues, four of which participate in disulfide bonds (Cys177-Cys236 and Cys368-Cys415) and one of which is free (Cys246). In this report, we investigated the mechanism of thermal inactivation of variant 3C6P and its *H. jecorina* Cel6A and *Humicola insolens* Cel6A parents by exploring the effects of removing free cysteines.

3.3 Methods

This section describes the methods used to study thermal inactivation in this Chapter. For materials and general methods, please refer to Chapter 6.

3.3.1 Residual activity measurements

Depending on the pH of the reaction, sodium acetate buffer (pH 5) or sodium phosphate buffer (pH 6 – 9) was used. Samples containing 5 μg of purified Cel6A in 40 μL of 50 mM buffer were inactivated at 70°C to 90°C for 15 minutes for residual activity measurements. After thermal inactivation, 60 μL of 5% (w/v) Avicel in double-deionized water was added and incubated at 50°C for 2 hours. Alternatively, samples containing 2 μg of purified Cel6A in 40 μL of 50 mM buffer were inactivated at 85°C for 15 minutes. The samples were cooled to 4°C for 45 minutes before addition of 60 μL of 5% (w/v) Avicel in double-deionized water and incubation at 50°C for 2 hours. The supernatant was analyzed using Nelson-Somogyi reducing sugar assay with cellobiose as the reducing sugar standard. Residual activity data were modeled with the 4-parameter Boltzmann sigmoidal equation ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$).

3.3.2 Residual activity in the presence of Avicel

Depending on the pH of the reaction, sodium acetate buffer (pH 5) or sodium phosphate buffer (pH 6 – 9) was used. Samples containing 15 μg of purified Cel6A and 3% (w/v) Avicel in 100 μL of 20 mM buffer were inactivated at 70°C to 90°C for 15 minutes for residual activity measurements. After thermal inactivation, the enzymes were incubated with Avicel at 4°C for 1 hour to promote maximum

substrate binding. The sugar produced during the inactivation period was subsequently removed by washing the Avicel-bound enzyme twice with 180 μL of fresh 20 mM buffer and resuspending the solids in 75 μL of 20 mM buffer. The Avicel-bound enzymes were incubated at 50°C for 2 hours to determine residual activity. The supernatant was analyzed using Nelson-Somogyi reducing sugar assay with cellobiose as the reducing sugar standard. Residual activity data were modeled with the 4-parameter Boltzmann sigmoidal equation ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$).

3.3.3 Half-life measurements

Depending on the pH of the reaction, sodium acetate buffer (pH 5) or sodium phosphate buffer (pH 6 – 9) was used. Samples containing 5 μg of Cel6A in 40 μL were inactivated at 90°C for up to 2 hours in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. After heat inactivation, the enzymes were incubated at 50°C for 2 hours with 60 μL of 5% w/v Avicel in double deionized water to determine the residual activity. Half-lives were determined from plots of the natural logs of residual activities vs. the inactivation time.

3.3.4 Circular dichroism measurements

CD measurements were collected on an AVIV 62DS spectrometer equipped with a thermoelectric cell holder using a 1 mm path length cell. Spectra were recorded from 250 nm to 200 nm at every 1 nm, with a bandwidth of 1 nm and an averaging time of 3 seconds. Three spectra were recorded at each temperature to derive an average signal and standard error of the measurements. The measurements were performed at pH 5 – 9 in 10 mM buffer (sodium acetate at pH 5 and sodium phosphate at pH 6 – 9). The enzyme concentration in the CD experiment was 10 μM. The CD signals were corrected with calibration curves from 2.4 mM of (D)-(-)-pantolactone (Sigma-Aldrich) and 2.3 mM of (+)-camphor-10-sulfonic acid (Sigma-Aldrich) according to a published protocol [102].

3.3.5 Crystallographic image processing

The FFT functionality under Maps & Mask Utilities in CCP4 suite [73] was used to convert the MTZ files corresponding to the deposited structures of HJPlus and 3C6P to the $2F_o$-F_c map and the F_o-F_c map, where F_o and F_c refer to observed and calculated structure factors. All protein structure images were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). The $2F_o$-F_c map was displayed using isomesh at 1σ level, while the F_o-F_c map was displayed using isomesh at 3σ level to indicate missing electron densities and -3σ level to indicate excess electron densities.

3.4 Results

3.4.1 Effect of removing free cysteine in engineered thermostable Cel6A

Thermostable Cel6A variant 3C6P is optimally active at 75°C but inactivates irreversibly within 5 minutes at 90°C [88]. A Cel6 sequence alignment revealed near-complete conservation of the two disulfide bridges (202 out of 205 protein sequences examined), whereas the remaining free cysteine Cys246 is not (55% Gly, 20% Cys). We introduced a glycine at position 246 in 3C6P and measured residual activity after thermal inactivation at 70 – 90°C, pH 5 – 9. At pH 5, the residual activity of 3C6P decreases sigmoidally with increasing temperature, and the enzyme was completely inactivated after incubation above 85°C (Figure 3.1A); the temperature of inactivation decreased with increasing pH. Variant 3C6P C246G was also inactivated with increasing temperature, but never lost all activity after the temperature was reduced. 3C6P C246G retained some activity even after incubation at 90°C at all pH (Figure 3.1B).

The half-life of 3C6P C246G was examined at 90°C, pH 5 – 9, conditions under which 3C6P is completely inactivated within five minutes. The half-lives varied from 8 to 76 minutes (Table 3.1); 3C6P C246G is most stable at pH 6. The C246G mutation significantly increases the ability of this cellulase to tolerate extreme temperatures.

To better understand the effect of the C246G mutation, we examined whether the change in the
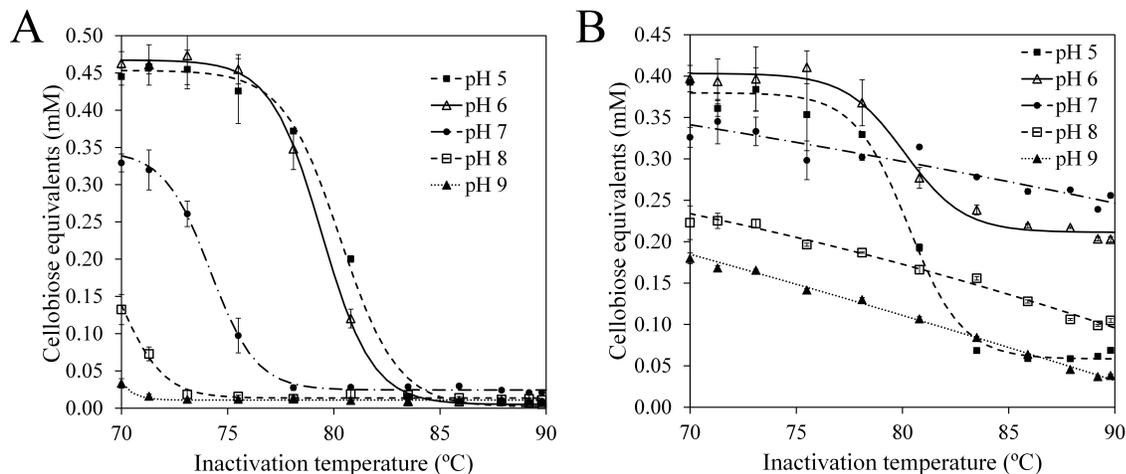
Figure 3.1: Residual activity profiles of 3C6P (A) and 3C6P C246G (B). Purified enzymes were inactivated at different temperatures for 15 minutes at pH 5 – 9 before they were assayed for activity at 50°C for 2 hours with 3% (w/v) Avicel. Activity is reported as the total cellobiose equivalents released.

thermal inactivation profile (Figure 3.1) was due to the absence of cysteine or the presence of glycine.

The Cel6 alignment revealed at least four other amino acids at this position: A (13%), L (9%), S (2%), and N (1%). These mutations were introduced into 3C6P. Variant 3C6P C246N did not express in our *S. cerevisiae* expression host, and no further experiments were performed for it. Residual activities were determined after thermal inactivation at 70 – 90°C and pH 7, where the difference in residual activity profile is the most striking. As shown in Figure 3.2, all four variants with mutations at residue C246 have residual activity profiles similar to that of 3C6P C246G, suggesting that the absence of cysteine is important for retention of catalytic activity after heat treatment.

Table 3.1: Half-lives of 3C6P C246G at 90°C and pH 5 – 9. Purified enzyme was incubated at 90°C for different lengths of time and then assayed for activity at 50°C with 3% (w/v) Avicel.

Reaction pH	$t_{1/2, 90°C}$ (min)
pH 5	8.3 ± 4.3
pH 6	76 ± 10
pH 7	48 ± 5
pH 8	20 ± 2
pH 9	8.9 ± 1.7
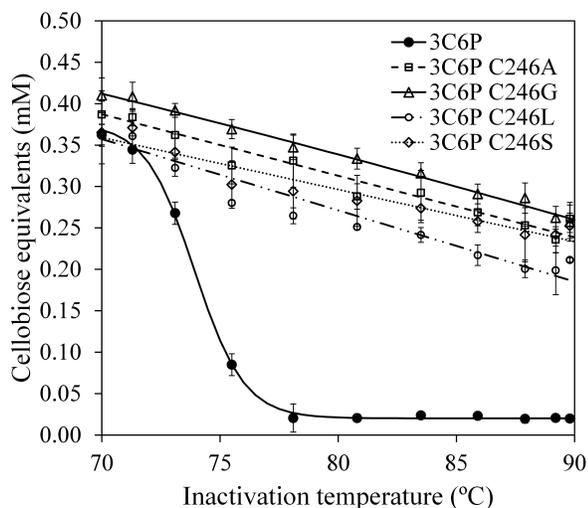
Figure 3.2: Effects of different mutations at residue C246 on the residual activity of 3C6P. Purified enzymes were inactivated at different temperatures for 15 minutes before they were assayed for activity at 50°C for 2 hours with 3% (w/v) Avicel in 20 mM sodium phosphate buffer, pH 7. Activity is reported as the total cellobiose equivalents released.

Variant 3C6P is from a family of chimeric Cel6A cellobiohydrolases that have either one or two free cysteines, C246 and C400. We sought to determine whether a free cysteine at residue 400 (which is a serine in 3C6P) alters residual activity and also whether both the highly conserved disulfide bonds contribute to the ability of 3C6P C246G to withstand high temperature incubation. Four variants of 3C6P C246G were thus constructed: 3C6P C246G having the S400C mutation, 3C6P C246G having only one disulfide bond (CC1, CC2), and 3C6P C246G with no disulfide bonds (CC0). Table 3.2 summarizes the residual activities after thermal inactivation at 85°C and pH 6, the pH at which 3C6P C246G has the longest half-life. Mutation S400C reduces the residual activity of 3C6P C246G to the level of 3C6P, indicating that the absence of free cysteines is necessary for high thermotolerance. In addition, removing the disulfide-forming cysteine pair Cys177-236 reduced the residual activity of 3C6P C246G by one-third, while the absence of Cys368-Cys415 renders the thermal inactivation almost fully irreversible, whether Cys177-236 is present or not.

Table 3.2: Residual activities of cysteine and disulfide variants of 3C6P. Purified enzymes were heat-inactivated at 85°C for 15 minutes and assayed for activity at 50°C with 3% (w/v) Avicel in 20 mM sodium phosphate buffer, pH 6. % residual activity is reported relative to activity before heat inactivation.

Variant	\multicolumn{4}{c}{Residue position}	% residual activity
	246	400	177-236	368-415	
3C6P	C	S	C-C	C-C	1.6 ± 0.2
3C6P C246G	G	S	C-C	C-C	67 ± 1
3C6P C246G S400C	G	C	C-C	C-C	3.4 ± 0.8
CC1	G	S	C-C	S-S	8.7 ± 1.1
CC2	G	S	S-S	C-C	44 ± 2
CC0	G	S	S-S	S-S	7.2 ± 0.6

3.4.2   Thermal denaturation by CD spectroscopy

To better understand the thermotolerance conferred by the C246G mutation, we compared the secondary structures of 3C6P and 3C6P C246G using circular dichroism (CD) spectroscopy at pH 5 – 9. At each pH, wavelength scans were recorded at 25°C to establish the spectrum of the folded enzyme and at 90°C to observe the secondary structure at high temperature. Wavelength scans were conducted again once the sample was cooled to 25°C.

Figure 3.3 shows the CD spectra of 3C6P and 3C6P C246G, grouped by the temperatures at which the measurements were taken. Both variants have very similar spectra at 25°C over the pH range 5 – 9, with a global minimum at 208 nm and a local minimum around 225 nm. At 90°C, both 3C6P and 3C6P C246G retain significant secondary structure. The global minimum shifted to 204 nm for both proteins, indicating a large fraction of random coil. When 3C6P was cooled to 25°C, the global minimum of the 3C6P spectra remained at 204-206 nm and approximated the spectra at 90°C, indicating irreversible unfolding. In contrast, the spectra of 3C6P C246G after heat treatment and cooling at pH 6 – 9 follow the native spectra closely, with a global minimum at 208 nm and a local minimum at 225 nm. Thermal denaturation of 3C6P C246G thus appears to be partially reversible at pH 6 – 9, which is in agreement with the residual activity data (Figure 3.1B). The CD spectra of 3C6P C246G after thermal denaturation are very similar at pH 7 – 9, yet the enzyme
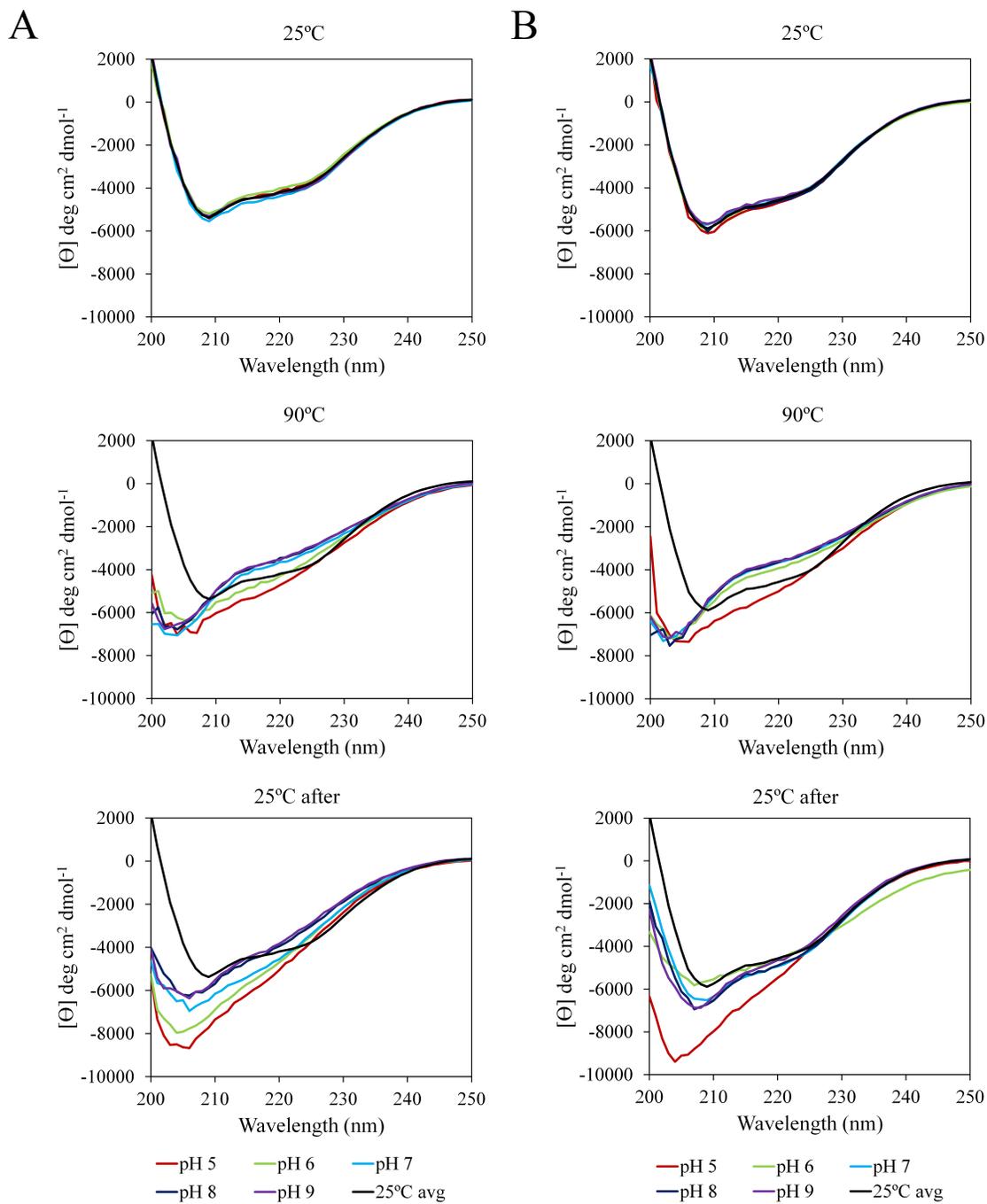
Figure 3.3: Far-UV wavelength scans of 3C6P (A) and 3C6P C246G (B) with CD spectroscopy. CD spectra were taken at 25°C, 90°C, and 25°C again after heat treatment. The average scans of 3C6P or 3C6P C246G at 25°C are shown in black for comparison.

retains different levels of activity, implying that retention of native-like structure is not sufficient to recover native activities. Other factors such as deamidation of asparagine and glutamine, whose rate increases with increasing pH [103], may play a role as well. Methionine has also been reported to undergo oxidation at high temperatures [89]. At pH 5, the spectrum of 3C6P C246G at 25°C after thermal denaturation retained the global minimum at 204 nm, indicating random coil in the structure and minimal refolding consistent with loss of activity (Figure 3.3).

3.4.3   Residual activities in the presence of Avicel

To determine whether the observed residual activity profile is unique to enzyme inactivation in the absence of substrate, heat inactivation of 3C6P and 3C6P C246G were conducted in the presence of Avicel. Variant 3C6P is observed to withstand inactivation to higher temperatures in the presence of Avicel (Figure 3.4), as the inactivation temperatures that 3C6P lost all activities are higher in the presence of Avicel than in the absence. Instead of a sharp transition between the active and the inactivated state, the residual activity of 3C6P declined linearly rather than sigmoidally with increasing temperature in the presence of Avicel. However, 3C6P is still completely inactivated at 90°C. On the other hand, similar to the residual profile in the absence of Avicel, 3C6P C246G retained residual activity at pH 6 – 9 after incubation at 90°C. At pH 5, it is difficult to ascertain whether 3C6P C246G retained residual activity after thermal inactivation at 90°C as the cellobiose equivalents released were too close to the detection limit of the Nelson-Somogyi assay.

3.4.4   Residual activities in the presence of DTT

To investigate how the disulfide bonds contribute to retention of activity after heat treatment, variants 3C6P and 3C6P C246G were inactivated at 70 – 90°C and pH 5 – 9 in the presence of 1 mM dithiothreitol (DTT). While DTT reduces disulfide bonds readily at pH 7 or above, it has also been observed to retain its reducing power at pH 4.5 or 5 at high temperatures [104, 105]. As shown in Figure 3.5, the residual activities of 3C6P and 3C6P C246G do not differ significantly in the presence of DTT. They do differ, however, from the profiles measured without DTT (Figure 3.1).
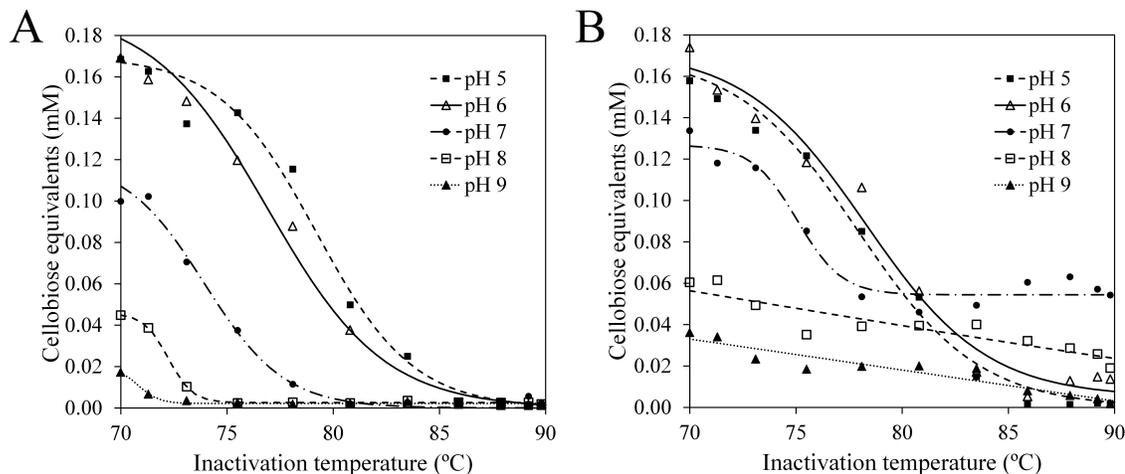
Figure 3.4: Residual activity profiles of 3C6P (A) and 3C6P C246G (B) inactivated in the presence of Avicel. Purified enzymes were inactivated in the presence of 3% (w/v) Avicel at different temperatures for 15 minutes at pH 5 – 9. The sugar produced during inactivation was washed away before the enzymes were assayed for activity at 50°C for 2 hours. Activity is reported as the total cellobiose equivalents released.

The activities of 3C6P C246G are higher when DTT is not present during incubation, indicating that the reversibility of thermal denaturation depends on correct formation of the disulfide bonds.

3.4.5   Crystallographic observations

In Chapter 2, crystal structure of 3C6P was determined at 1.2 Å (PDB ID 4I5U) to study the mechanism behind the thermostabilizing mutations [88]. The protein crystal was obtained in sodium acetate buffer at pH 5.8. During crystal structure construction, when the conserved cysteine pair Cys177-Cys236 was modeled as cystine, the electron density between the cysteine side chains was continuous, indicating that the residues were participating in a disulfide bond. However, the F_o-F_c difference map showed a strong negative peak between the sulfur atoms, indicating that the disulfide configuration did not represent the structure adequately and that the cysteines existed in both oxidized and reduced states in the protein crystal (Figure 3.6). In addition, Cys236 could be modeled with a reduced rotamer pointing toward the carbonyl oxygen of Asn226 and the carbonyl oxygen of Asn230. A similar observation was made when modeling Cys368-Cys415 as well, although the effect is much more pronounced in Cys177-Cys236 than Cys368-Cys415 (Figure 3.6).
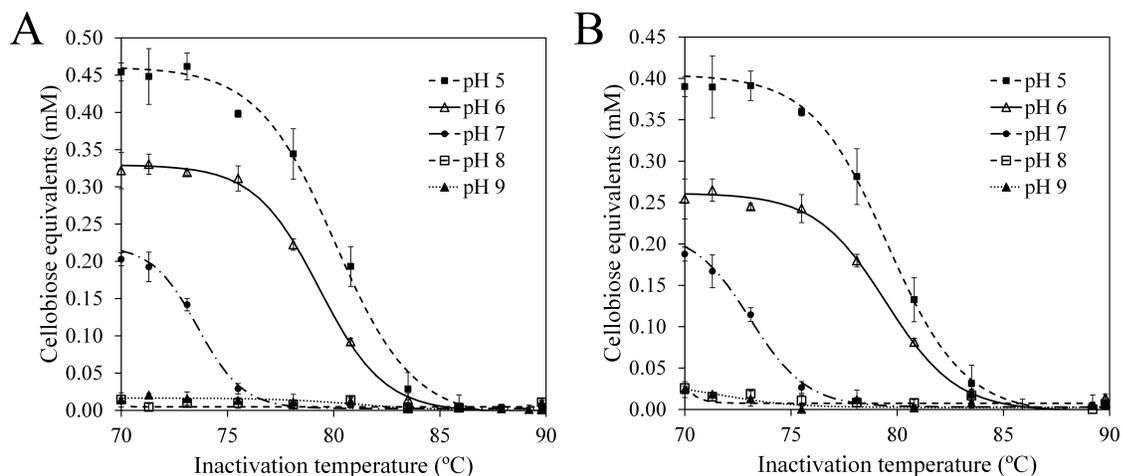
Figure 3.5: Residual activity profiles of 3C6P (A) and 3C6P C246G (B) inactivated in the presence of reducing agent DTT. Purified enzymes were inactivated at different temperatures at pH 5 – 9 before they were assayed for activity at 50°C with 3% (w/v) Avicel, 1 mM DTT.

3.4.6 Effect of removing free cysteines in *H. jecorina* and *H. insolens* Cel6A

We examined whether the thermotolerance gained by removing free cysteines is unique to the engineered thermostable Cel6A or whether there is a similar benefit in wild-type Cel6A. Cel6A from *H. insolens* (HiCel6A) and *H. jecorina* (HjCel6A) contain one (Cys403) and two (Cys245 and Cys400) free cysteines, respectively. These were mutated to serines, and the residual activities of the variants were determined after thermal inactivation at 85°C, pH 6. Both variants containing no free cysteine
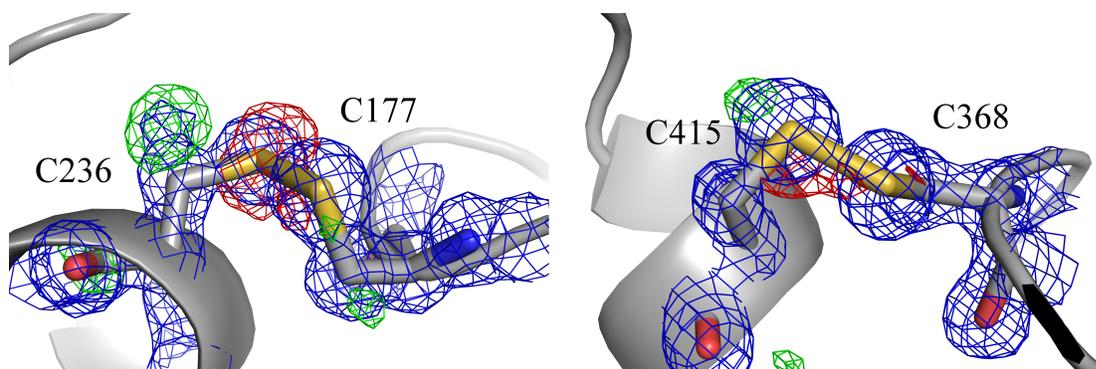
Figure 3.6: Cysteine pairs Cys177-Cys236 and Cys368-Cys415 in 3C6P (PDB ID 4I5U) in both oxidized and reduced conformations. Blue represents $2F_o$-F_c electron density map displayed at 1σ level. Green represents missing electron density from F_o-F_c difference map displayed at 3σ level, and red represents excess electron density from F_o-F_c difference map displayed at -3σ level.

Table 3.3: Residual activities of wild-type and cysteine variants of *H. jecorina* (Hj) and *H. insolens* (Hi) Cel6A. Purified enzymes were heat-inactivated at 85°C for 15 minutes and assayed for activity at 50°C with 3% (w/v) Avicel in 20 mM sodium phosphate buffer, pH 6. % residual activity is reported relative to activity before heat inactivation.

Variant	Residue position (HiCel6A/HjCel6A)		% residual activity
	249/245	403/400	
HiCel6A	L	C	1.6 ± 0.2
Hi C403S	L	S	42 ± 2
HjCel6A	C	C	3.2 ± 0.3
Hj C245S C400S	S	S	70 ± 6

exhibited significant activity after thermal inactivation, whereas HiCel6A and HjCel6A were fully inactivated (Table 3.3). After heat treatment, HjCel6A lacking its free cysteines recovered 70% activity, comparable to 3C6P C246G under the same conditions, whereas the HiCel6A variant lacking its free cysteine retained 42% activity. Removing the free cysteines thus allows the wild-type Cel6A to better tolerate incubation at high temperatures.

3.5 Discussion

Thermal inactivation is the irreversible loss of activity when an enzyme is held at high temperatures for a prolonged period of time [106]. A better understanding of the mechanisms of thermal inactivation can aid us in engineering enzymes for industrial applications. In the present study, we investigated the impact of free cysteines on irreversible thermal inactivation of Cel6A. All four 3C6P mutants lacking Cys246 (C246A, C246G, C246L, and C246S) retained some activity after 15-minute incubation at 90°C and pH 7, in contrast to 3C6P, which lost all activity. CD spectroscopy and thermal inactivation with DTT indicate that the residual activity of 3C6P C246G stems from disulfide-bond-assisted refolding to a native-like conformation. The mechanism of thermal inactivation of Cel6A appears to involve disulfide bond degradation and thiol-disulfide exchange at high temperatures. Degradation of disulfide bonds is initiated when hydroxyl ions either attack the sulfur atom or deprotonate the α-carbon (β-elimination) or the β-carbon of the cysteine (α-elimination)

[93]. Though believed to occur mainly in neutral or basic pH, β-elimination of disulfide bonds has been observed at pH as low as pH 4 in a number of proteins at 100°C [91, 107]. The half-life of disulfide bonds at 100°C is much longer at pH 6 than pH 8, which is in agreement with the half-lives of 3C6P C246G at 90°C reported here.

Thiol-disulfide exchange refers to the formation of new disulfide bonds when free thiols, including those generated by disulfide bond degradation, attack another disulfide bond to form a new cystine [108]. Such processes have been implicated in enzyme thermal inactivation at high temperatures. Zale and Klibanov demonstrated that the irreversible inactivation of ribonuclease, which does not have any free cysteines, was due to the shuffling of disulfide bonds at pH 6 and 8, and addition of free cysteines greatly increased the rate of inactivation at 90°C [91]. Similarly, the presence of the free cysteine in 3C6P significantly decreases the half-life at 85 – 90°C. In the absence of free cysteines, the rate of thermal inactivation is limited by disulfide bond degradation, which occurs two orders of magnitude more slowly than thiol-disulfide exchange [107].

The observation that the cysteines participating in disulfide bonds were found in both the bridged and reduced states in the protein crystal may be attributed to ionizing radiation damage from synchrotron X-ray. Ionizing radiation from synchrotron sources leaves specific damage to the protein structure, including disulfide bond cleavage and decarboxylation of acidic residues, instead of non-specific damage [109, 110]. In addition, there is a specific order of susceptibility to cleavage among the several disulfide bonds within a protein, and the order of susceptibility has been suggested to be related to the intrinsic stability of the different disulfide bonds [110]. In the crystal structure of 3C6P, the disulfide bond Cys177-Cys236 is much more susceptible to cleavage than Cys368-Cys415, as indicated by the F_o-F_c difference maps. This coincides with the residual activity data of variant CC1 (containing only Cys177-Cys236) and variant CC2 (containing only Cys368-Cys415). In particular, variant CC2 retained an order of magnitude higher activity than variant CC1 after thermal inactivation at 85°C, confirming that Cys368-Cys415 is more resistant to degradation than Cys177-Cys236.

Two wild-type Cel6A related to 3C6P, *H. insolens* Cel6A (HiCel6A) and *H. jecorina* Cel6A

(HjCel6A), are both irreversibly inactivated at temperatures above 70°C, whereas Cel6A from *Chaetomium thermophilum* (CtCel6A) partially refolds when the temperature is reduced [88]. We hypothesize that this behavior reflects the presence or absence of free cysteines in the respective enzymes. Variants of HiCel6A and HjCel6A having no free cysteines retained up to 70% activity after thermal inactivation at 85°C, which suggests that wild-type Cel6A follow the same mechanism of thermoinactivation as the thermostable 3C6P. We postulate that Cel6A enzymes have a propensity to refold to their native conformation after thermal denaturation. However, free cysteines increase the rate of thermal inactivation by participating in thiol-disulfide exchange and causing the protein to misfold and aggregate.

In conclusion, fungal cellobiohydrolases from family 6 glycoside hydrolase are essential components of cellulases used to degrade cellulose to produce higher-value fuels and chemicals. Understanding the mechanism of thermal inactivation in this enzyme family can aid future efforts to engineer them to withstand harsh industrial conditions. Thermal inactivation involves disulfide-bond degradation and thiol-disulfide exchange in engineered thermostable Cel6A as well as the wild-type enzymes. Removal of free cysteines is essential if these enzymes are to tolerate high temperatures. The high residual activities of Cel6A variants lacking free cysteine stems from disulfide-bond-assisted refolding to native-like structure after thermal inactivation.

Chapter 4

Thermostability and Ionic Liquid Tolerance

4.1 Abstract

Lignocellulose is a complex substrate that is recalcitrant to enzymatic degradation, and substrate pretreatments are necessary to increase the accessibility of lignocellulose to cellulases and hemicellulases. Ionic liquids like 1-ethyl-3-methylimidazolium (EMIM) acetate are emerging as promising pretreatment chemicals, but the production and recycling of ionic liquids are energy-intensive and costly to implement. A cost-effective method is to eliminate the need to separate ionic liquids before enzyme hydrolysis and use cellulases and hemicellulases that are tolerant to ionic liquids. In this chapter, we compared different thermostability measurements to assess the thermostability of fungal Cel6A cellobiohydrolases and established correlations between thermostability measurements and EMIM acetate tolerance. The inactivation effects of EMIM acetate are more severe at higher temperatures, and the variants tolerated lower EMIM acetate concentrations at 70°C than 50°C. The presence of EMIM acetate lowered the activities of the Cel6A variants, with the exception that 5% EMIM acetate slightly enhanced the activity instead. The enhancing effect was observed for all variants with an IC_{50} of 10% or higher at 50°C and/or 70°C. Future investigations are needed to verify that the correlations hold between thermostability and EMIM acetate tolerance with lignocellulosic substrates and for other families of cellulases.

4.2 Introduction

Lignocellulose consists mainly of three types of biopolymers — cellulose, hemicellulose, and lignin — that are tightly woven together and connected by covalent bonds and non-covalent interactions [111]. Composed of long linear chains of glucose, cellulose is the main component of lignocellulose and forms extensive intermolecular hydrogen bonds with other cellulose chains and hemicellulose [28]. Hemicelluloses are branched heterogeneous polymers of pentose (e.g. xylose) and hexose (e.g. mannose, glucose), and the exact composition of the hemicellulose is dependent on the biomass source [112]. Cellulose and hemicellulose like xylan can be converted to their respective monomeric sugars by cellulases and xylanases. However, the tight knitted structure of lignocellulose and the presence of lignin make the substrate very recalcitrant to enzymatic degradation [111, 113].

Different physical and chemical pretreatment conditions have been applied to increase the accessibility of lignocellulose to cellulases and xylanases by separating cellulose, hemicellulose, and lignin from each other. Dilute sulfuric acid pretreatment, steam explosion, and ammonia fiber expansion (AFEX) are all commonly employed as pretreatment methods in the production of lignocellulosic biofuels [11]. Each pretreatment method has its own advantages and drawbacks, and the high cost of many pretreatment chemicals calls for their recovery and recycling. The choice of an optimal pretreatment method requires the careful balance between minimizing the production of inhibitory compounds, reducing lignin content, and the safe handling of the pretreatment chemicals, while maximizing the accessibility of lignocellulose [13].

Ionic liquids (IL) are low melting-point salts that are emerging as an alternative pretreatment chemical. A number of studies have examined the effects of different IL in pretreating lignocellulose, and the combinations of imidazolium-based or pyridinium-based cations and acetate or chloride as the anion are the most effective in dissolving cellulose [18]. In addition, the ability of 1-alkyl-3-methylimidazolium-based IL to dissolve cellulose increases with decreasing alkyl chain length. Even-numbered alkyl chains are also better at dissolving cellulose than odd-numbered alkyl chains [18]. After dissolution of cellulose, an anti-solvent such as water is added to precipitate cellulose and separate lignin [114]. Pretreatment with 1-ethyl-3-methylimidazolium (EMIM) acetate has been

shown to significantly decrease the crystallinity of cellulose and reduce lignin content in switchgrass, which translate to faster hydrolysis rates and higher total conversion [115, 116]. However, application of IL pretreatment is currently limited by the high cost of the ionic liquid, and recovering ionic liquid after pretreatment is energy-intensive and costly to implement [18].

An alternative cost-effective approach is to eliminate the separation step of ionic liquid after pretreatment [116]. This requires that the cellulases and hemicellulases for hydrolysis be highly tolerant of the ionic liquid used for pretreatment. However, the activity of commercial cellulase mixtures is strongly attenuated in the presence of 10% EMIM acetate at 50°C [117, 118]. Studies have reported the discovery of IL-tolerant cellulases from (hyper)thermophilic bacteria and archaea and demonstrated that more thermostable cellulases retain activity better in the presence of ionic liquid [119, 120]. Previously, we created a family of fungal Cel6A cellobiohydrolases by structure-guided recombination [45]. Chimera 12222332, or HJPlus, is one of the chimeras closest in sequence to the industrially relevant Cel6A from *Hypocrea jecorina*. By conducting random mutagenesis and recombination of beneficial mutations on HJPlus, we obtained a series of thermostable Cel6A cellobiohydrolases [88]. In this chapter, we evaluated different thermostability measurements for assessing the Cel6A variants and identified correlations between EMIM acetate tolerance and thermostability in fungal Cel6A cellobiohydrolases. The result demonstrates that increasing thermostability is an effective strategy for increasing EMIM acetate tolerance.

4.3 Methods

This section describes the methods used in this chapter to study thermostability and ionic liquid tolerance. For materials and general methods, please refer to Chapter 6.

4.3.1 Thermostability measurements

4.3.1.1 T_{50} measurements

All T_{50} measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. Samples containing 2 μg of Cel6A in 40 μL were inactivated at different temperatures for 15 minutes in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. The temperature range was selected to ensure the T_{50} value is bracketed by a 20°C range. After heat inactivation, the enzymes were incubated at 50°C for 2 hours with 60 μL of 5 % w/v Avicel to measure the enzymes residual activity. To determine T_{50}, the residual activities were plotted against the temperature using SigmaPlot (Systat Software Inc) and fitted using the 4-parameter Botlzmann sigmoidal function ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$). T_{50} is the inactivation temperature with half maximal residual activity, or x_0.

4.3.1.2 Melting temperature measurement by DSF

SYPRO Orange indicator dye (Invitrogen) was diluted to 10X concentration in 1.5 mL of 100 mM HEPES buffer, pH 7.4, with 150 mM sodium chloride. Samples containing 15 μL of 10 μM purified Cel6A were combined with 15 μL of 10X SYPRO Orange and subjected to thermal melting from 25°C to 99°C at a rate of 1°C/min. The fluorescence signals at excitation wavelength of 490 nm and emission wavelength of 530 nm were monitored every 30 seconds using a StepOne Plus Real-time PCR machine (Applied Biosystems). To determine the melting temperature (T_m), the fluorescence signals were plotted against the temperature using GraphPad Prism (GraphPad software) and fitted with the 4-parameter Boltzmann sigmoidal function ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$). T_m is the temperature with half maximal fluorescence intensity, or x_0.

4.3.1.3 T_{A50} measurements

All T_{A50} measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. Samples containing 2 μg of purified Cel6A in 40 μL were combined with 60 μL of 50 mg/mL Avicel and incubated at different temperatures in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. The incubation temperatures were selected to include the optimum temperature (temperature with the

highest activity) and temperatures at which the enzyme loses all activity after 2 hours. The activity was determined by measuring the reducing sugar concentration in the reaction supernatant using Nelson-Somogyi sugar assay. To determine T_{A50}, the activity data was first normalized with the optimum activity, which was obtained by averaging the cellobiose concentrations of the three highest activity readings. The temperature at which the activity first reached 95% or above was set as the cut-off point, and activity data at temperatures below the cut-off point was discarded. The remaining activity data was plotted against the temperature using SigmaPlot (Systat Software Inc) and fitted with the 4-parameter Boltzmann sigmoidal function ($y = y_0 + \frac{a}{1+exp^{-(x-x0)/b}}$). T_{A50} is the temperature with half maximal activity, or x_0.

4.3.2 Activity in the presence of ionic liquid

All activity assays with ionic liquid were conducted with 50 mM sodium acetate buffer, pH 5.0. Samples containing 75 μL of 0.4 μM purified Cel6A were combined with 75 μL of 40 mg/mL Avicel or IL-pretreated switchgrass in 50 mM sodium acetate buffer supplemented with 10%, 20%, 30%, and 40% 1-ethyl-3-methylimidazolium (EMIM) acetate and incubated at 50°C or 70°C for 16 hours. The incubation was performed in a Mastercycler Pro Thermal Cycler (Eppendorf) with heated lid. Alternatively, samples containing 10 mg of IL-pretreated switchgrass were suspended in 125 μL of 10%, 20%, 30%, and 40% 1-ethyl-3-methylimidazolium (EMIM) acetate and 61.5 μL of 50 mM sodium acetate buffer and incubated with 13.5 μL of 40 mg/mL bovine serum albumin (BSA) at 50°C or 70°C for 1 hour. After the pre-incubation, 50 μL of 4.25 μM purified 3C6P were combined with ILSG and incubated at 50°C or 70°C for 16 hours. The incubation was performed in a tabletop Thermomixer Comfort (Eppendorf) shaking at 1400 rpm. After the 16-hour incubation, the reducing sugar concentration in the reaction supernatant was determined using dinitrosalicylic acid (DNS) sugar assay.

4.4 Results and discussion

4.4.1 Stabilizing mutations from mutagenesis libraries

Upon mutagenesis of HJPlus, we found five positions (M135, Q277, S317, S406, and S413) with stabilizing mutations that increased the thermostability of the Cel6A enzymes [88]. Saturation mutagenesis libraries with NNK degeneracy were constructed at each position in HJPlus and screened for total activity at 75°C after 2 hours. For each NNK library, 90 colonies were examined, a roughly three-fold oversampling. For the NNK libraries at all positions, the beneficial mutations identified from the random mutagenesis libraries were also found as the top variants in the NNK libraries. For position S317 and S413, two new beneficial substitutions were identified: S317W and S413W. Table 4.1 summarizes the T_{50} measurements of the Cel6A variants with different combinations of stabilizing mutations. T_{50} is defined as the temperature at which the enzyme loses 50% of its activity after 15 minutes of thermal inactivation. In general, the stabilizing mutations discovered as single mutations were found to be stabilizing in combination with other mutations, with the exception of the mutations at residues 317 and 413. The most thermostable variant has prolines at both positions, even though tryptophan has been shown to be comparably stabilizing as single mutants in either position.

4.4.2 Comparison between T_{50} and other thermostability measurements

The twelve Cel6A variants in Table 4.1 cover a range of 8.2°C in T_{50} measurements. We sought to evaluate the effectiveness of different thermostability measurements by comparing T_{50} with other measures. The melting temperature of the enzyme based on loss of structural elements has been commonly used to evaluate thermostability. The melting temperatures (T_m) of the variants were determined with differential scanning fluorimetry (DSF) using the indicator SYPRO Orange, whose fluorescence is unquenched when it binds to hydrophobic regions as the enzyme unfolds [121, 122]. As shown in Figure 4.1A, a comparison between T_m and T_{50} showed no correlation between the two measurements ($R^2 = 0.03$). Melting temperatures determined by differential scanning calorimetry

Table 4.1: T_{50} measurements of the Cel6A variants. The amino acid at each position is displayed at the top of the column, while the mutated amino acids in each variant with respect to HJPlus are displayed in the grid.

Variant	S30	V128	M135	Q277	S317	S406	S413	T_{50} (°C)
HJPlus	-	-	-	-	-	-	-	71.9 ± 0.6
1G6	-	-	-	-	P	-	-	73.2 ± 0.3
317F12	-	-	-	-	W	-	-	73.6 ± 0.5
413A1	-	-	-	-	-	-	W	74.3 ± 0.3
1F4	-	-	-	-	-	-	F	73.0 ± 0.3
2B3	-	-	-	L	P	-	-	75.7 ± 0.3
2F4	-	-	L	-	P	-	-	74.3 ± 1.1
2G6	-	-	-	-	P	P	-	74.9 ± 0.6
3C6	F	A	L	L	P	P	-	76.9 ± 0.2
3C6P	F	A	L	L	P	P	P	80.1 ± 0.4
3C4W	F	A	L	L	P	P	W	79.5 ± 0.4
3CW2	F	A	L	L	W	P	W	77.5 ± 0.5

or circular dichroism spectroscopy are needed to evaluate whether DSF is a valid method to assess the unfolding of Cel6A. In addition, the T_m measurements of the variants are consistently higher than the T_{50} measurements, suggesting that the variants are inactivated before the enzymes expose significant areas of hydrophobic regions.

The T_{50} measurements are based on residual activity data after incubation at high temperature. We sought to examine whether T_{50} correlates with activity at high temperatures. The optimum temperature of Cel6A, the temperature at which the catalytic activity is highest, is a likely candidate for examining activity at high temperatures. However, since the optimum temperatures of Cel6A tend to cover a broad range (~ 5°C), a different measurement of thermostability, T_{A50}, was investigated. T_{A50} is defined as the temperature at which the enzyme loses 50% activity relative to activity at its optimum temperature. As shown in Figure 4.1B, a comparison between T_{A50} and T_{50} measurements of the variants showed a good correlation between the two ($R^2 = 0.90$), unlike the relationship between T_m and T_{50}. T_{A50} is consistently higher than T_{50} for all the variants examined, which is likely due to the fact that Avicel exhibits stabilizing effects with cellulases. In the presence
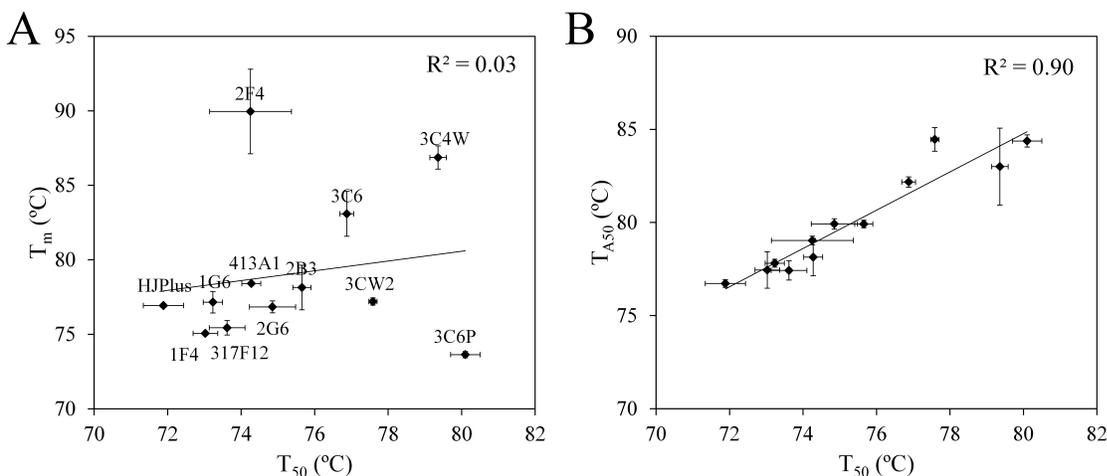
Figure 4.1: Comparison between different thermostability measurements. A) The correlation between T_{50} and T_m. B) The correlation between T_{50} and T_{A50}. T_{50} is the temperature at which the enzyme loses 50% of its activity after 15-min heat inactivation. T_m is the apparent melting temperature determined by monitoring the thermal unfolding of Cel6A with SYPRO Orange. T_{A50} is the temperature at half-maximal activity. All three thermostability measurements were obtained by modeling the activity or fluorescence data with the 4-parameter Boltzmann sigmoidal function.

of Avicel, the enzymes undergo thermal denaturation at a slower rate due to their interaction with the substrate.

4.4.3 IL tolerance with Avicel as substrate

To determine if a correlation exists between Cel6A thermostability and tolerance to ionic liquid, the activities of the variants were measured at 50°C or 70°C for 16 hours in the presence of different concentrations of EMIM acetate using Avicel as substrate. To quantify EMIM acetate tolerance, IC_{50} of the variants were determined by fitting the activity at different concentrations of ionic liquid with a 4-parameter sigmoidal plot. IC_{50} is defined as the inhibitor concentration at which the enzyme loses 50% of maximal activity. At 50°C, IC_{50} did not vary dramatically (from 15% to 18%) among the variants examined, though a reasonable correlation between IC_{50} and T_{50} at 50°C (R^2 = 0.83, Figure 4.2A) existed. At 70°C, the variation in IC_{50} is larger (from 5% to 11%), and a stronger correlation exists between IC_{50} and T_{50} at 70°C (R^2 = 0.91, Figure 4.2A). The correlations improved further when IC_{50} was compared with T_{A50}, with a R^2 of 0.84 at 50°C and R^2 of 0.94 at 70°C (Figure 4.2B). At both temperatures, the tolerance to EMIM acetate, IC_{50}, increases as the
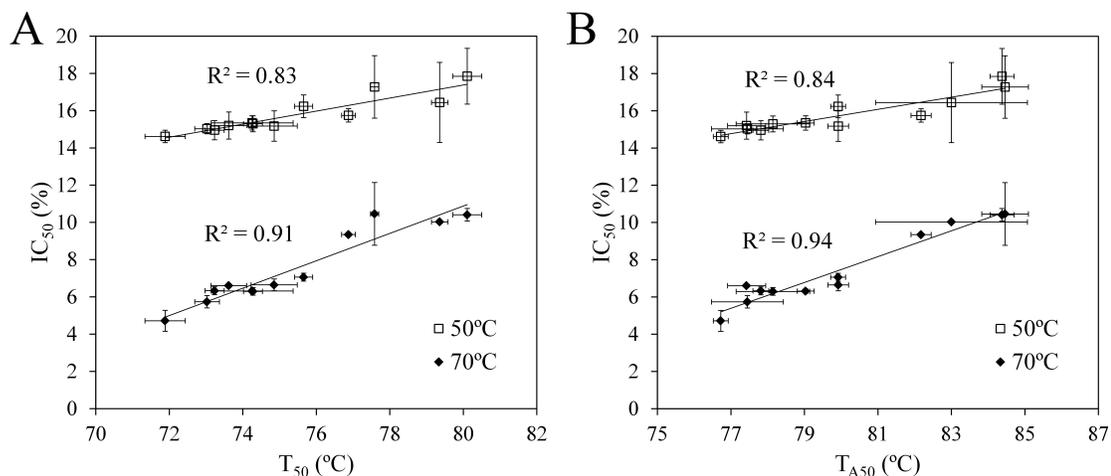
Figure 4.2: Correlation between thermostability and ionic liquid tolerance at 50°C and 70°C. A) Correlation between T_{50} and IC_{50}. B) Correlation between T_{A50} and IC_{50}. Activity in the presence of EMIM acetate was determined by incubating Cel6A with 3% (w/v) Avicel and 0 – 20% EMIM acetate at 50°C or 70°C for 16 hours. IC_{50} is the EMIM acetate concentration at half-maximal activity determined by modeling the activity data with the 4-parameter Boltzmann sigmoidal function.

thermostability of the variants increases. In all the variants examined, IC_{50} at 50°C is higher than IC_{50} at 70°C, indicating that the activity declined faster at 70°C than at 50°C as the concentration of EMIM acetate increased. In other words, the effects of thermal inactivation and inactivation by EMIM acetate are cumulative, and EMIM acetate is more detrimental to cellulase activity at a higher temperature.

4.4.4 IL tolerance with pretreated biomass as substrate

Avicel is a commercially available substrate of microcrystalline cellulose commonly used to determine the activities of cellobiohydrolases, but it does not adequately reflect the complexity of lignocellulosic substrate used in biofuel production. We sought to determine if similar trends exist between T_{50} and ionic liquid tolerance when lignocellulosic substrate is used. The activities of the variants were again determined at 50°C or 70°C for 16 hours in the presence of different concentrations of EMIM acetate using IL-pretreated switchgrass (ILSG). ILSG has been washed and dried to remove any ionic liquid from the substrate pretreatment process, and the different concentrations of EMIM acetate were achieved by suspending the pretreated biomass in sodium acetate buffer supplemented

with EMIM acetate. Unlike Avicel, which is uniform in size with a particle size of ~ 50 μm, ILSG is not homogeneous in size, has a much larger average particle size, and is difficult to pipet in a liquid suspension. As a result, the well-to-well variation in the amount of ILSG dispensed was large. A new substrate dispensing method is needed to improve the well-to-well variation, either by accurately dispensing milligrams of dry substrate per well or by dispensing liquid substrate dissolved in ionic liquid. However, the amount of ILSG in each well was in excess compared to Cel6A and should not be the limiting factor in this hydrolysis experiment. Unfortunately, the activities of the variants on ILSG after 16 hours were below the detection range of DNS reducing sugar assay. This is most likely due to the lignin content in the substrate, to which cellulases bind non-specifically and non-productively [29].

4.4.5 IL tolerance in the presence of BSA

To obtain sufficient cellulase activity on IL-pretreated switchgrass (ILSG), it was necessary to either include an additive to block non-specific binding to lignin or use a different lignocellulosic substrate with a lower lignin content before the experiment can proceed. Different surfactants and additives have been added during cellulose hydrolysis to decrease nonspecific lignin binding [123, 124]. In particular, studies by the Wyman group have shown that addition of bovine serum albumin (BSA) increases the hydrolysis yield by preventing non-productive binding of cellulase onto lignin and decreasing the crystallinity of cellulose [125, 126]. We sought to determine if the addition of BSA, a non-catalytic protein in the mixture, would allow us to measure cellulase activity on ILSG in the presence of EMIM acetate. Due to the difficulty in dispensing liquid-suspended ILSG, in this experiment 10 mg of ILSG were carefully weighed and dispensed in individual sample tubes to ensure that the tube-to-tube variation was at a minimum. BSA was pre-incubated with ILSG prior to the addition of 3C6P. As shown in Figure 4.3, the activity of 3C6P on ILSG at 50°C was similar as the concentration of EMIM acetate increased from 10% to 20%, suggesting that the combination of ILSG and BSA dampens the inactivating effect of EMIM acetate at 50°C. However, the same trend was not observed with the activity on ILSG at 70°C. Furthermore, the activity in the presence
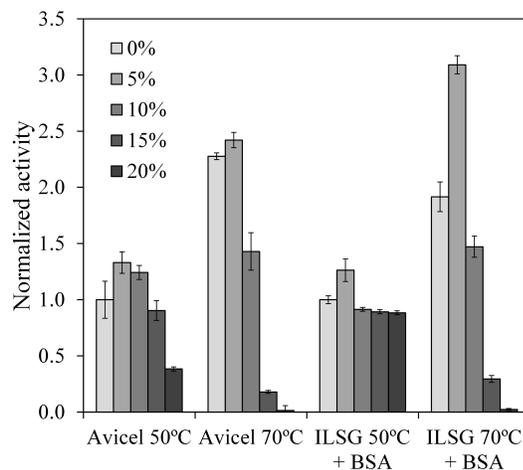
Figure 4.3: Normalized activity of variant 3C6P on Avicel and IL-pretreated switchgrass with different concentrations of EMIM acetate. Purified 3C6P was incubated at 50°C or 70°C for 16 hours with 3% (w/v) Avicel or 4% (w/v) ILSG supplemented with 0 – 20% EMIM acetate. The ILSG samples were pre-incubated with 7.5% BSA w/w of ILSG for 1 hour before the addition of 3C6P. Activity on Avicel was normalized by the activity at 50°C and 0% EMIM acetate on Avicel, while activity on ILSG was normalized by the activity at 50°C and 0% EMIM acetate on ILSG.

of EMIM acetate would suggest that the IC_{50} at 50°C is higher with ILSG than Avicel but that the IC_{50} at 70°C is higher with Avicel than ILSG. Further investigation is needed to elucidate if different cellulose substrates alter the IC_{50} of the Cel6A variants. Also, the presence of BSA likely alters the dynamics of the enzyme denaturation and further complicates the quest to understand the effect of ionic liquid. As a result, a lignocellulosic substrate with a lower lignin content than the present study should be used instead of supplementing the hydrolysis reaction with BSA.

4.4.6 Activity enhancing effect of 5% EMIM acetate

The Cel6A activity on Avicel decreases as the concentration of EMIM acetate increases. The exact mechanism of inactivation — let it be denaturation, inhibition to catalysis, or inhibition to binding — requires further investigation. Interestingly, 5% of EMIM acetate slightly enhanced the activity at 50°C on Avicel for most Cel6A variants examined in this study. The enhancing effect of 5% EMIM acetate increased the Cel6A activity at 50°C by up to 38%, and it does not correlate with the thermostability of the variants (Figure 4.4). At 70°C, the enhancing effect is only observed for variants with IC_{50} above 10%, increasing the activity of those variants by up to 16% (Figure 4.4).
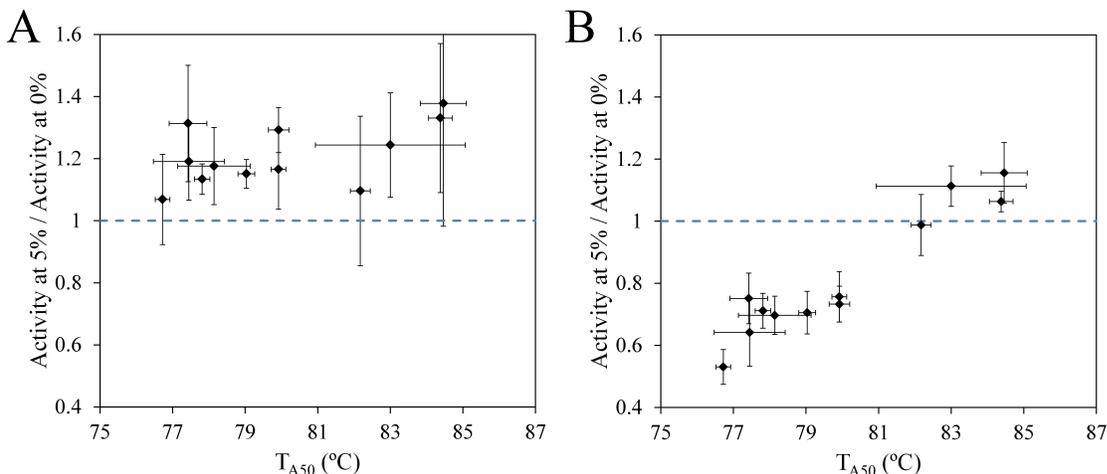
Figure 4.4: Effect of 5% EMIM acetate on Cel6A activity on Avicel. A) Activity ratio between 5% and 0% EMIM acetate at 50°C. B) Activity ratio between 5% and 0% EMIM acetate at 70°C. Activity was determined by incubating Cel6A with 3% (w/v) Avicel supplemented with 0% or 5% EMIM acetate at 50°C or 70°C for 16 hours. The dashed line represents when the activity in the presence of 0% and 5% are equivalent, above which 5% EMIM acetate has an enhancing effect on the activity of Cel6A.

The cutoff of IC_{50} at 10% holds true at 50°C as well, as all Cel6A variants have IC_{50} above 14% at this temperature.

Similarly, 5% EMIM acetate also enhanced the activity of 3C6P on both Avicel and ILSG at 50°C and 70°C (Figure 4.3), and the effect is most pronounced on ILSG at 70°C, increasing the activity of 3C6P by 61%. Further investigation is needed to study the enhancing effect with activity on ILSG and ensure that BSA is not the cause of the activity enhancement.

4.5   Future directions

This chapter details the comparison between different thermostability measurements and ionic liquid tolerance as determined by IC_{50}. Although the correlation between T_{A50} (or T_{50}) and IC_{50} is excellent with a R^2 value as high as 0.94, the thermostability range covered by the Cel6A variants in this study is limited to ~8°C in T_{A50} or T_{50}. Inclusion of additional Cel6A variants with thermostability outside of this range, such as wild-type Cel6A from *H. jecorina* (HjCel6A) and Cel6A chimeras constructed with HjCel6A, would significantly increase the applicability of the established

correlation. In addition, examining whether the same correlation holds true with other family and types of fungal cellulases would also increase the impact of the study.

An efficient cellulase mixture includes different endoglucanases and cellobiohydrolases as well as other accessory enzymes that work synergistically to hydrolyze lignocellulosic substrate. Consequently, the optimal cellulase for hydrolysis in the presence of ionic liquid would not only be tolerant to ionic liquid but also retain its synergism with other cellulases. To demonstrate the advantage of using thermostable cellulases as ionic-liquid tolerant enzymes, a long-time hydrolysis experiment should be conducted using thermostable mixtures of cellulases to hydrolyze lignocellulosic substrate in the presence of ionic liquid as compared to wild-type cellulase mixtures or commercial cellulase mixtures.

Chapter 5

Semi-rational Protein Design

5.1 Abstract

Small libraries or rationally designed mutations are desirable for protein engineering as evaluating large mutagenesis libraries can be time consuming and energy intensive. Several semi-rational protein designs were evaluated in this chapter. Consensus sequence analysis, backbone stabilization by prolines, and B-factor analysis were used to design mutations to enhance the thermostability of variant 3C6P. However, none of the consensus amino acid substitutions or designed proline mutations increased the T_{50} measurements of 3C6P, though it might be necessary to verify the contribution of the neutral mutations by examining their activity at high temperatures. A correlation was established between the change in the free energy of unfolding estimated using FoldX ($\Delta\Delta G_{FoldX}$) and the change in T_{50} (ΔT_{50}). Even though the mutations predicted to be stabilizing by $\Delta\Delta G_{FoldX}$ did not enhance the T_{50} of 3C6P, the correlation between $\Delta\Delta G_{FoldX}$ and ΔT_{50} demonstrated that $\Delta\Delta G_{FoldX}$ can be used to eliminate point mutations that have detrimental effects to thermostabilitfy. Lastly, saturation mutagenesis libraries designed on the basis of high average B-factors identified several mutations with higher total activity than 3C6P in medium-throughput

Cel6A screens at 80°C. Detailed characterizations of the mutants are needed to understand the contribution of these mutations. Small libraries or rationally designed mutations are desirable for protein engineering as evaluating large mutagenesis libraries can be time consuming and energy intensive. Several semi-rational protein designs were evaluated in this chapter. Consensus sequence analysis, backbone stabilization by prolines, and B-factor analysis were used to design mutations to enhance the thermostability of variant 3C6P. However, none of the consensus amino acid substitutions or designed proline mutations increased the T_{50} measurements of 3C6P, though it might be necessary to verify the contribution of the neutral mutations by examining their activity at high temperatures. A correlation was established between the change in the free energy of unfolding estimated using FoldX ($\Delta\Delta G_{FoldX}$) and the change in T_{50} (ΔT_{50}). Even though the mutations predicted to be stabilizing by $\Delta\Delta G_{FoldX}$ did not enhance the T_{50} of 3C6P, the correlation between $\Delta\Delta G_{FoldX}$ and ΔT_{50} demonstrated that $\Delta\Delta G_{FoldX}$ can be used to eliminate point mutations that have detrimental effects on thermostability. Lastly, saturation mutagenesis libraries designed on the basis of high average B-factors identified several mutations with higher total activity than 3C6P in medium-throughput Cel6A screens at 80°C. Detailed characterizations of the mutants are needed to understand the contribution of these mutations.

5.2 Introduction

Random mutagenesis is an effective strategy for evolving a protein of interest for improvements in thermostability, novel activity, and substrate specificity [127]. However, evaluating large libraries of variants is time consuming and energy intensive. The frequency of beneficial mutation is low, and a large fraction of the random mutagenesis library is often rendered inactive by deleterious mutations [128, 129]. The trend in directed evolution has been moving toward constructing smaller and smarter libraries that enrich for active variants [130]. The advances in the mechanistic understanding of proteins and the increasing availability of protein structures have also prompted rational protein designs as the emerging choice for protein engineering [131].

In this chapter, several semi-rational protein engineering approaches were applied to further

enhance the thermostability of 3C6P, the most thermostable Cel6A obtained thus far. The mutations were chosen on the basis of consensus sequence analysis, backbone stabilization by prolines, and B-factor analysis. Consensus-guided mutagenesis has been applied to increase the thermostability of various enzymes. The hypothesis is that since stabilizing mutations are more likely to be kept during protein evolution, consensus amino acids from homologous proteins have a higher probability of contributing positively to protein stability than nonconsensus amino acids [132, 133, 86]. Another method of increasing thermostability is to decrease the entropy of the unfolded state by introducing proline substitutions that restrict backbone conformation [134]. Studies have reported that prolines at the N-cap positions of helices and in β-turns were stabilizing [135]. The stabilizing proline substitutions in 3C6P were also found in or near loop regions [88]. As a result, proline substitutions were designed to replace residues residing at the N-cap positions of -helices and 3_{10} helices as well as residues at the i+1 positions of Type I, II, and IV β-turns. The last semi-rational approach is saturation mutagenesis on the basis of high B-factors, which has been shown to dramatically increase the T_{50}^{60} of *Bacillus subtilis* lipase by 45°C [136]. B-factors in crystal structures represent atomic smearing of electrons relative to their equilibrium positions, and residues with high B-factors are perceived to have large movements and thus represent flexible regions within the protein [137]. By targeting residues with high B-factors using saturation mutagenesis, we sought to discover amino acid substitutions that provide rigidity and enhance the thermostability of 3C6P.

5.3   Methods

This section describes semi-rational protein engineering approaches used in this chapter. For materials and general methods, please refer to Chapter 6.

5.3.1   Multiple sequence alignments

Cel6A protein sequence was analyzed using pBLAST by searching the query sequence against the non-redundant protein sequence database. Pairwise alignments were run between each sequence from the database and the query sequence using the Multiple Sequence Comparison by Log-Expectation

(MUSCLE) tool. Sequences having less than 25% identity with the query sequence or sequences that were composed of more than 50% gaps were disregarded. The consensus sequence is defined as the sequence in which more than 50% of the aligned sequences share the same amino acid at a particular position. The identity of the amino acid was left blank when there was no amino acid that was represented by more than 50% of the aligned sequences at a particular position.

5.3.2 FoldX energy calculation

FoldX version 3.0 beta4 was used for all the energy calculation. The structure of 3C6P (PDB ID 4I5U) was first repaired by removing the water molecules and ligands and moving the side chains to remove bad torsion angles and Van der Waals clashes. The mutations, either the point proline mutations or the position scans, were generated as semicolon separated lists and analyzed using the BuildModel functionality. The difference in the energy of unfolding between the mutant and 3C6P was reported with the unit in kcal/mol.

5.4 Results

5.4.1 Consensus amino acid substitutions

Cel6A from *Humicola insolens*, *Hypocrea jecorina*, and *Chaetomium thermophilum* were used to identify Cel6 sequences that share 25 – 95% sequence identity with the query sequences. After eliminating sequences with more than 50% gaps from the query sequences, a total of 205 Cel6 sequences were aligned and analyzed for consensus sequence. A sequence alignment between 3C6P and the consensus sequence as well as the list of all the Cel6 sequences used in the multiple sequence alignment can be found in Appendix D. Eighteen positions were found to have a strong consensus amino acid (over 50%) whose identity is different from the amino acid in 3C6P. The consensus amino acid substitutions were constructed in 3C6P, with the exception of mutation M135V as it coincides with the beneficial mutation M135L found in the random mutagenesis library. An NNK saturation mutagenesis library at position 135 identified leucine as the most stabilizing mutation

(Chapter 4). The T_{50} measurements of the consensus mutations are listed in Table 5.1, with the amino acids representing 51 – 66% consensus at the indicated positions. However, none of the consensus mutations increased the T_{50} of 3C6P, suggesting that these mutations are either neutral or destabilizing.

5.4.2   Designed proline substitutions

Three out of five stabilizing mutations in 3C6P are serine-to-proline substitutions in loop regions. A crystal structure alignment between HJPlus and 3C6P indicates that the loop conformations are conserved by the proline substitutions [88]. Secondary structural analysis of the 3C6P crystal

Table 5.1: T_{50} measurements of the strong consensus mutations in 3C6P. Multiple sequence alignment of 205 Cel6 sequences indicate eighteen strong consensus mutations, which are substitutions with amino acids that are present in 50% of the Cel6 sequences at the indicated positions.

Enzyme variant	Consensus (%)	T_{50} (°C)
3C6P	-	80.1 ± 0.4
3C6P W99Y	63	79.2 ± 0.2
3C6P N102P	66	77.2 ± 0.2
3C6P R122A	51	78.7 ± 0.4
3C6P A124K	51	75.4 ± 0.3
3C6P M146L	52	78.8 ± 0.4
3C6P I153A	66	79.4 ± 0.4
3C6P Y186L	62	80.0 ± 0.3
3C6P C246G	55	80.1 ± 0.6
3C6P V251L	58	80.1 ± 0.3
3C6P S292G	57	79.4 ± 0.2
3C6P L297V	52	79.3 ± 0.5
3C6P P321W	51	76.8 ± 0.5
3C6P F334L	53	79.3 ± 0.2
3C6P P358G	53	69.2 ± 1.1
3C6P G360R	53	74.7 ± 0.4
3C6P Q361R	64	76.3 ± 0.2
3C6P T373A	61	76.6 ± 0.2

structure was performed using PROMOTIF [138] as part of the PDBSum analysis functionality [139]. Two of the proline mutations, S317P and S413P, reside on the N-terminal caps of 3_{10} helices, while the third serine-to-proline mutation, S406P, resides on the i+1 position of Type I β-turn. As a result, additional proline substitutions were designed to mimic the proline substitutions from the random mutagenesis libraries.

To stabilize helices, we chose three proline substitutions that reside on the N-cap position of 3_{10} helices and four proline substitutions that reside on the N-cap position of -helices. The pyrrolidine ring of proline restricts the number of conformations that the amino acid can adopt [140]. All the chosen residue positions have ϕ angles near -60° and ψ angles near -35° with the exception of one (-17°), which are angles that proline can tolerate. The proline substitutions were introduced in 3C6P, and the thermostability of the proline mutants were evaluated using T_{50} measurements. However, none of the proline substitutions as N-caps of helices were stabilizing (Table 5.2).

To stabilize the loop regions, we chose ten residues based on their locations as the i+1 residue in β-turns for substitution with proline in 3C6P. All chosen residue positions again have ϕ angles near -65° and ψ angles near -20° or 130°, which are also angles that prolines can tolerate. In addition, none of the chosen residues forms a hydrogen bond at its backbone amide, and most lie in exposed loops. Based on the T_{50} measurements of the proline mutants, these proline mutations are either neutral or destabilizing (Table 5.3). To test whether the proline substitutions might stabilize a less thermostable Cel6A, four proline mutations corresponding to mostly neutral proline mutations in 3C6P (3C6P S211P, 3C6P L231P, 3C6P A381P, and 3C6P E386P) were introduced in the Cel6A from *H. insolens* (HiCel6A). These proline mutants have T_{50} values that are comparable to HiCel6A (Table 5.4), suggesting that the mutations are again neutral with the exception of Q390P.

Strategic incorporation of prolines in loop regions stabilizes the protein by restricting the conformational freedom of the peptide backbone [135, 89, 141]. Given that prolines have the highest propensity to reside in the i+1 position of Type I and II β-turns [142], we examined fourteen proline substitutions in β-turns in 3C6P and HiCel6A. These proline substitutions are all neutral or destabilizing with the exception of one. One possible explanation is that stabilizing prolines by

Table 5.2: T_{50} measurements of 3C6P with proline mutations residing on the N-cap positions of α-helices or 3_{10} helices. The $\Delta\Delta G_{FoldX}$ is the change in the free energy of unfolding predicted by FoldX relative to the free energy of 3C6P.

Enzyme variant	Helix	$\Delta\Delta G_{FoldX}$ (kcal/mol)	T_{50} (°C)
3C6P	-	-	80.1 ± 0.4
3C6P N102P	α-helix	-0.4	77.2 ± 0.2
3C6P I113P	3_{10} helix	5.5	70.1 ± 0.7
3C6P L140P	3_{10} helix	-0.6	77.6 ± 0.1
3C6P I188P	3_{10} helix	1.7	76.1 ± 0.2
3C6P L224P	α-helix	3.3	71.2 ± 0.1
3C6P A268P	α-helix	3.9	70.3 ± 0.1
3C6P Q432P	α-helix	-1.9	76.4 ± 0.3

Table 5.3: T_{50} measurements of 3C6P with proline mutations residing in β-turns. The $\Delta\Delta G_{FoldX}$ is the change in the free energy of unfolding predicted by FoldX relative to the free energy of 3C6P.

Enzyme variant	β-turn	$\Delta\Delta G_{FoldX}$ (kcal/mol)	T_{50} (°C)
3C6P	-	-	80.1 ± 0.4
3C6P S211P	Type I	-0.3	79.6 ± 0.2
3C6P L231P	Type I	0.7	78.2 ± 0.4
3C6P V304P	Type II	2.2	69.7 ± 0.2
3C6P G352P	Type IV	6.1	72.1 ± 0.4
3C6P W364P	Type I	1.4	72.3 ± 0.1
3C6P A381P	Type IV	-1.4	79.9 ± 0.2
3C6P E386P	Type I	-1.0	80.3 ± 0.3
3C6P V394P	Type IV	5.1	75.6 ± 0.2
3C6P G397P	Type II	-0.0	77.9 ± 0.3
3C6P A427P	Type II	-1.6	76.9 ± 0.3

Table 5.4: T_{50} measurements of *H. insolens* Cel6A (HiCel6A) with proline mutations residing in β-turns.

Enzyme variant	β-turn	T_{50} (°C)
HiCel6A	-	65.2 ± 0.1
HiCel6A S215P	Type I	65.6 ± 0.3
HiCel6A M235P	Type I	63.1 ± 0.1
HiCel6A A385P	Type IV	64.7 ± 0.2
HiCel6A Q390P	Type I	66.1 ± 0.4

rational incorporation in β-turns has been reported in mostly small proteins (18 kDa or smaller) [135, 141, 143, 144, 145, 146, 147]. Similar proline substitutions in a larger protein are either neutral or destabilizing in the same study [147]. Combined with our results, this suggests that prolines are well tolerated in β-turns if placed strategically, but the incorporation of prolines in loops does not necessarily confer greater stability.

5.4.3 FoldX analysis and predictions

FoldX is a protein design algorithm that uses energy functions with empirically determined coefficients to study the importance of atomic interactions in proteins. The effects of mutations on protein stability can be estimated by comparing the free energy of unfolding between variants [148]. The $\Delta\Delta G_{FoldX}$ was computed for seventeen proline mutants compared to the free energy of unfolding for 3C6P and are listed in Tables 5.2 and 5.3. Based on the $\Delta\Delta G_{FoldX}$ calculation, FoldX estimated that seven out of seventeen prolines have a negative $\Delta\Delta G_{FoldX}$ value, suggesting that these mutations are stabilizing. Even though none of the proline mutations increased the T_{50} of 3C6P, a correlation between $\Delta\Delta G_{FoldX}$ and T_{50} was found with a R^2 of 0.56 (Figure 5.1). Qualitatively speaking, the seven proline mutants with negative $\Delta\Delta G_{FoldX}$ values all had T_{50} within -4°C from the T_{50} of 3C6P, while the proline mutants with positive $\Delta\Delta G_{FoldX}$ values had T_{50} at least 4°C lower than the T_{50} of 3C6P, with the exception of one (L231P). In other words, $\Delta\Delta G_{FoldX}$ might be useful to qualitatively compare the effects of different mutations but cannot be used to predict
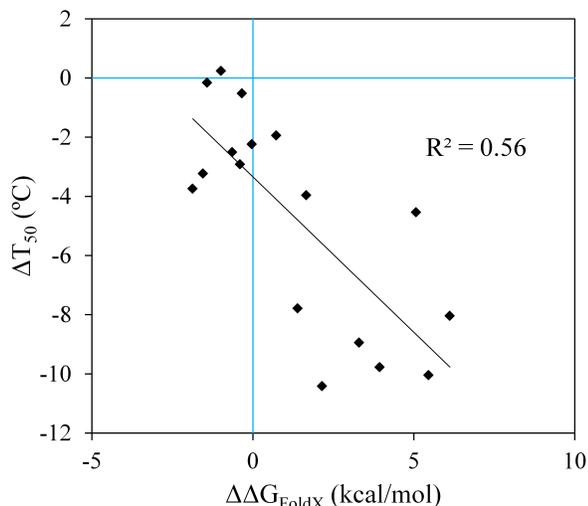
Figure 5.1: Correlation between ΔT_{50} and $\Delta\Delta G_{FoldX}$ of the proline mutants. The $\Delta\Delta G_{FoldX}$ is the change in the free energy of unfolding predicted by FoldX relative to the free energy of 3C6P, while ΔT_{50} is the change in T_{50} measurement relative to the T_{50} of 3C6P. A negative $\Delta\Delta G_{FoldX}$ indicates that the mutation is predicted to be stabilizing, while a positive $\Delta\Delta G_{FoldX}$ indicates that the mutation is predicted to be destabilizing. A negative ΔT_{50} indicates that the mutation is destabilizing.

the extent to which these mutations affect thermostability.

One of the proline mutations found in the random mutagenesis libraries, S317P, is a proline in the wild-type Cel6A from *H. insolens*, one of the parents used to construct the Cel6A chimera library. A close examination of the Cel6A from *H. insolens* and from *C. thermophilum* indicated that four prolines exist in the wild-type Cel6A that are not found in 3C6P. In particular, two of the proline mutations are adjacent insertion mutations that are truncated in 3C6P and the Cel6A from *H. jecorina* (Figure 5.2). The effects of the insertion proline mutation (163PP) and two wild-type prolines (D344P and S445P) in the thermostability of 3C6P were evaluated. FoldX was used to estimate the $\Delta\Delta G_{FoldX}$ of point mutations and predicted both wild-type proline substitutions to be stabilizing. In addition, FoldX also predicted that a weak consensus proline substitution (A235P, 43% Pro) to be stabilizing, which was constructed in 3C6P and evaluated as well. All four variants had T_{50} measurements comparable or lower than 3C6P (Table 5.5).

One possible explanation that $\Delta\Delta G_{FoldX}$ does not accurately predict stabilizing mutation is that the change in the energy of unfolding of the above mutations is not large enough compared
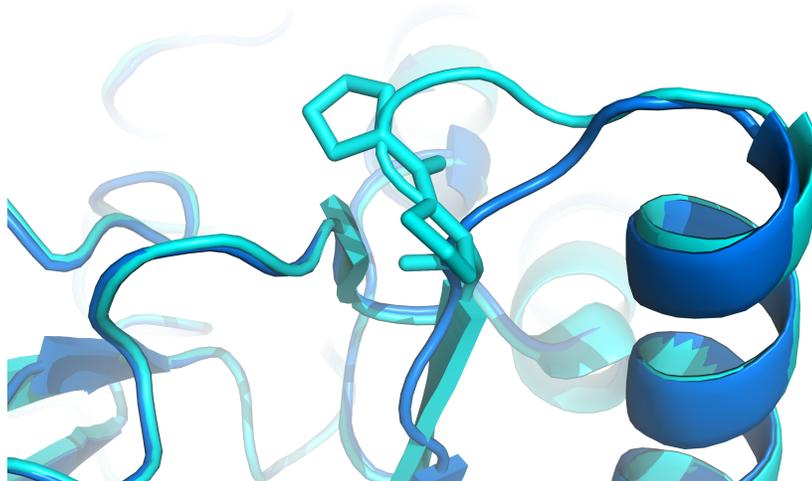
Figure 5.2: The truncated proline loop in 3C6P (PDB 4I5U, purple) compared to *C. thermophilum* Cel6A (PDB 4A05, cyan). The prolines missing in 3C6P are shown as sticks in Cel6A. The structure alignment was performed in Pymol.

to the error in the calculation. As a result, the energy of unfolding for all possible mutations in 3C6P — 19 amino acid substitutions for all 358 residues in the catalytic domain — were determined. The seven mutations listed in Table 5.6 have $\Delta\Delta G_{FoldX}$ less than -3, which are lower than all the proline mutations examined thus far. Since the reversion mutations (L135M, L277Q, P317S, P406S, and P413S) of the stabilizing substitutions in 3C6P all have $\Delta\Delta G_{FoldX}$ of ~2 with the exception of L135M, it might be of merit to examine the mutations with $\Delta\Delta G_{FoldX}$ around -3 or -4 for enhancement in thermostability.

Table 5.5: T_{50} measurements of 3C6P with proline mutations from wild-type Cel6 sequences. The $\Delta\Delta G_{FoldX}$ is the change in the free energy of unfolding predicted by FoldX relative to the free energy of 3C6P. The $\Delta\Delta G_{FoldX}$ calculation was not performed for 3C6P 163PP.

Enzyme variant	Type of mutation	$\Delta\Delta G_{FoldX}$ (kcal/mol)	T_{50} (°C)
3C6P	-	-	80.1 ± 0.4
3C6P 163PP	Insertion	ND	79.8 ± 0.3
3C6P D344P	Substitution	-0.8	78.8 ± 0.3
3C6P S445P	Substitution	-1.2	79.2 ± 0.4
3C6P A225P	Consensus	-2.3	80.1 ± 0.3

Table 5.6: Amino acid substitutions predicted to be stabilizing with $\Delta\Delta G_{FoldX}$ lower than -3 kcal/mol. The $\Delta\Delta G_{FoldX}$ is the change in the free energy of unfolding predicted by FoldX relative to the free energy of 3C6P.

Mutations	$\Delta\Delta G_{FoldX}$ (kcal/mol)
E107Y	-3.2
E107L	-3.1
S133L	-3.2
Q166L	-3.5
Q166M	-3.0
D171L	-4.4
D171N	-3.5

5.4.4 Saturation mutagenesis based on B-factors

The B-factors for all the atoms in HJPlus and 3C6P were extracted from the PDB structures (PDB 4I5R and 4I5U), and the average B-factors for each residue were calculated. As shown in Figure 5.3, the highest average B-factors for HJPlus are near positions 94, 118, 181, and 407, while the highest average B-factors for 3C6P are near positions 94, 118, 386, and 409. The average B-factors for HJPlus tend to be higher than that for 3C6P due to the fact that HJPlus is a lower resolution structure than 3C6P [136]. The structures of HJPlus and 3C6P have a backbone root mean square deviation (RMSD) of 0.24 Å, but the residues with the highest average B-factors from the structure of HJPlus and from the structure of 3C6P are different, suggesting that the choice of structure for B-factor analysis is crucial to the library design.

Saturation mutagenesis libraries with NNK degeneracy were constructed for residues Glu94, Asp118, Glu386, and Pro409 in 3C6P and screened for total activity at 80°C in 96-well plates. For each NNK library, 90 variants were examined, a roughly three-fold of oversampling. A total of thirteen variants exhibiting 10% or higher activity than 3C6P from all four libraries were selected for rescreening. However, none of the variants had total activities that were 10% higher than 3C6P during the rescreen. Given that the four positions selected for single NNK libraries also had nearby residues with high average B-factors, additional single and double NNK mutagenesis libraries were
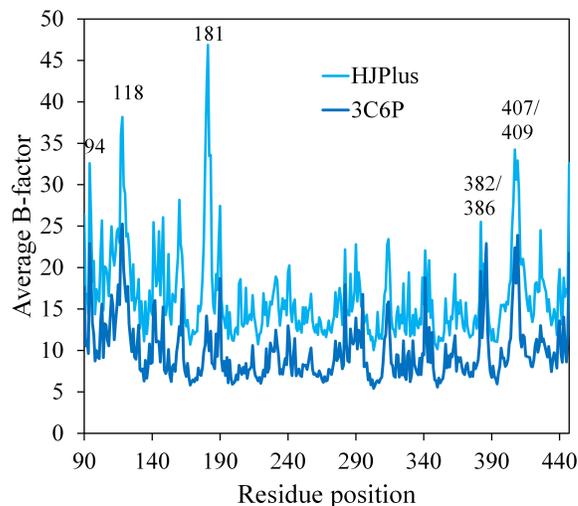
Figure 5.3: Average B-factors for the residues in HJPlus (PDB 4I5R) and in 3C6P (PDB 4I5U). The labeled positions are the residues with the highest average B-factors in HJPlus and/or 3C6P.

constructed at Thr117-Asp118, Asn382, His385-Glu386, and Ala408-Pro409 and screened for total activity at 80°C in 96-well plates. A total of 90 and 450 variants were sampled for the single and double NNK libraries respectively, and thirteen variants exhibiting 15% or higher activity than 3C6P from all four libraries were selected for rescreening. Five variants had a total activity that was 10% or higher than 3C6P during the rescreen. DNA sequencing revealed that the mutations were N382A, N382D, N382G, A408N-P409P, and A408S-P409P. Interestingly, the two improved mutants from the Ala408-Pro409 double NNK library both contain proline at position 409, indicating that a single NNK library at Ala408 might be a more suitable library design. Since the improvement in total activity at 80°C could stem from an improvement in expression, catalytic efficiency, or thermostability, detailed characterization of the mutants using purified enzymes is needed to evaluate how the mutations are beneficial.

The B-factor analysis identified several residues with high B-factors in close proximity of each other and provided many, perhaps too many, different library design options and combinatorial library designs. In addition, analysis done using different structures led to slightly different library designs, suggesting that the choice of protein structure is pivotal. As a result, using B-factor analysis as the basis of library designs might not be desirable if the goal is to limit the library size for screening.

5.5 Future directions

This chapter details the thermostability measurements of semi-rationally designed mutations using consensus sequence alignment, strategic incorporation of prolines, and B-factors from structure analysis. Even though none of the mutations is stabilizing in 3C6P, the FoldX analysis allowed us to identify a correlation between $\Delta\Delta G_{FoldX}$ and the change in T_{50}. In particular, the result demonstrates that $\Delta\Delta G_{FoldX}$ might be useful in avoiding mutations that are the most detrimental to the thermostability of 3C6P. In addition, seven mutations were identified as having $\Delta\Delta G_{FoldX}$ less than -3, which are promising substitutions for enhancing the thermostability of 3C6P.

The thermostability parameter used to evaluate the mutants is T_{50} in this chapter. However, as shown in Chapter 4, other thermostability parameters might be more suitable for the evaluation. Specifically, variant 3CW2 has a T_{A50} that is comparable to 3C6P but has a T_{50} that is -2.5°C lower than 3C6P. As a result, it is likely that many neutral mutations with comparable T_{50} to 3C6P (163PP, Y186L, A225P, S211P, V251L, A381P, and E386P) might have higher T_{A50} than 3C6P. Since T_{A50} closely reflects the activity at high temperatures, it is important to evaluate the T_{A50} measurements of these neutral mutants to determine if their contributions to thermostability are truly neutral.

Chapter 6

Materials and General Methods

6.1 Materials

6.1.1 Chemicals and commercial kits

For media preparation, yeast extract, peptone, Tryptone, and casamino acids were purchased from BD Bacto, and yeast nitrogen base without amino acids from BD Difco. SD-Ura powder was purchased from MP Biomedicals. LB powder and TB powder was purchased from RPI Corp. Sodium chloride, D-glucose, and carbenicillin were purchased from Sigma Aldrich. Sodium phosphate monobasic monohydrate and sodium phosphate dibasic heptahydrate were purchased from Mallinckrodt Chemicals.

For library construction, *Taq* DNA Polymerase was purchased from Roche. Phusion High-Fidelity DNA Polymerase, endonucleases (NheI, Acc65I, NdeI, KpnI, and NotI), and *Taq* DNA ligase were purchased from New England Biolabs. GeneMorph II Random Mutagenesis Kit was purchased from Stratagene. Sybr gold was purchased from Invitrogen. Pellet pain co-precipitant was purchased from Novogen.

For protein purification, Nelgene 0.2 μm PES filters, Vivaspin 20 concentrators with 10 kDa or 30 kDa MWCO, and imidazole were purchased from VWR. HisTrap HP columns, HiTrap Q HP columns, HiPrep 26/10 desalting columns, and PD-10 desalting columns were purchased from GE. Guanidine hydrochloride and quartz cuvettes were purchased from Sigma Aldrich.

For cellulase assays, Avicel PH-101, dithiothreitol (DTT), and ammonium molybdate were pur-

chased from Sigma Aldrich. 1-ethyl-3-methylimidazolium (EMIM) acetate and IL-pretreated switch-grass were provided by the Joint Bioenergy Institute.

QIAprep Miniprep Kit, QIAquick Gel Extraction Kit, and QIAquick PCR Purification Kit were purchased from Qiagen. Frozen-EZ Yeast Transformation II Kit and Zymoprep Yeast Plasmid Miniprep II Kit were purchased from Zymo Research.

6.1.2 Media preparation

For *E. coli*, LB medium was prepared by combining 5 g/L of yeast extract, 10 g/L of sodium chloride, and 10 g/L of Tryptone and sterilized by autoclaving. TB medium was prepared by dissolving 47.6 g/L of TB powder in double distilled, deionized water and sterilized by autoclaving. 2x YT medium was prepared by combining 10 g/L of yeast extract, 5 g/L of sodium chloride, and 16 g/L of Tryptone and adjusting the pH to 7.0 with sodium hydroxide (1 pellet of sodium hydroxide per liter). SOC medium was prepared by combining 5 g/L of yeast extract, 20 g/L of Tryptone, 0.58 g/L of sodium chloride, 0.19 g/L of potassium chloride, 0.95 g/L of magnesium chloride, and 3.6 g/L of D-glucose and sterilized by autoclaving.

For *S. cereviae*, SD-Ura medium was prepared by dissolving SD-Ura powder in double distilled, deionized water as directed and sterilized by autoclaving. SD-CAA medium was prepared by combining 8.6 g/L sodium phosphate monobasic monohydrate, 10.2 g/L sodium phosphate dibasic heptahydrate, 6.7 g/L yeast nitrogen base without amino acids, 5 g/L casamino acids, and 20 g/L D-glucose and sterilized by autoclaving. YPD medium was prepared by combining 10 g/L yeast extract, 20 g/L peptone, and 20 g/L D-glucose and sterilized by autoclaving.

All media sterilization by autoclaving was done at $120°C$ for 20 minutes.

6.2 Molecular cloning

6.2.1 Plasmid construction

All genes encoding Cel6A and Cel7A were cloned into yeast expression vector YEp352/PGK91-1-αss using NheI and KpnI endonuclease restriction sites and Gibson assembly. The wild-type and chimeric Cel6A genes were synthesized with *S. cerevisiae* codon bias and share the same N-terminal coding sequences for the cellulose binding module (CBM), the linker region, and the first five residues in the catalytic domain from the wild-type *H. jecorina* Cel6A. For protein expression in *E. coli*, the gene of interest was cloned into pET-22b(+) vector from Novogen using NdeI and NotI endonuclease restriction sites.

6.2.2 Primer designs

Primers used for error-prone PCR and overlap-extension PCR were designed using the following guidelines. Primers must begin (on the 5' end) and end (on the 3' end) with a G or C with $40-60\%$ GC content. If the beginning and the end of the primer cannot both contain a G or C, then designing a G or C on the 3' end takes priority over the 5' end. When the primer contains a mutation relative to the template, the mutation site is placed in the middle, bracketed by at least 12 bp of nucleotides on either side. The exact length of the flanking region is determined by the melting temperature of the region as determined by Vector NTI (Invitrogen), based on the Nearest Neighbor theory of DNA duplex stability [149]. The length of the flanking region is determined to achieve a melting temperature at $35-40°C$. The melting temperature of the overall primer should be $\sim60°C$ for a primer that is an exact match to the template and $70-75°C$ for a primer that contains a mutation relative to the template. The sequence used to insert His_6 tag is CATCACCATCACCACCAT, while the sequence used to insert His_8 tag is CATCACCATCACCACCATCACCAT. To introduce a new restriction site at the beginning (5' end) of the primer, a short nonsense sequence CTGCA is included prior to the restriction site to provide the endonucleases a handle for restriction digest.

6.2.3 Gibson assembly

To assemble gene insert with plasmid backbone, the DNA fragments need to share $50 - 75$ bp of overlaps. This was achieved by designing PCR primers $50 - 75$ bp upstream and downstream of endonuclease restriction sites. Plasmid backbone was constructed by digesting the expression vector with NheI and KpnI ($2 - 3$ U endonuclease per mg of plasmid) at $37°$C for 6 hours in NEBuffer 2 (1 mM Tris buffer, pH 7.9, supplemented with 5 mM sodium chloride, 1 mM magnesium chloride, and 0.1 mM DTT) and 0.1 mg/mL BSA. Gene insert was by constructed overlap-extension PCR, using 10 ng of plasmid, 20 nmol of dNTP, 25 pmol of forward primer and 25 pmol of reverse primer, 1x Phusion HF buffer, and 1 U of Phusion polymerase in a 50 μL reaction. The PCR program used 10 seconds at $98°$C for initial denaturation, with 30 cycles of 10 seconds at $98°$C, 10 seconds at $55°$C or $60°$C, and $72°$C for 30 seconds per 1000 nucleotides, with 5 minutes of final extension at $72°$C. The gene inserts were gel purified and extracted before assembling with plasmid backbone. Gibson master mix was made by mixing 320 μL of 5x isothermal reaction buffer with 6.4 μL of 1 U/μL T5 exonuclease, 20 μL of 2 U/μL Phusion polymerase, 160 μL of 40,000 U/μL *Taq* DNA ligase, and 694 μL of sterile deionized water. 5x isothermal reaction buffer was made by mixing 1.5 mL of 1 M Tris buffer, pH 7.5, with 75 μL of 2 M magnesium chloride, 120 μL of 400 mM dNTP, 150 μL of 1 M DTT, 0.75 g of PEG8000, 150 μL of 100 mM NAD, enough sterile deionized water to bring the final volume to 3 mL. To assemble the DNA fragments, 50 ng of backbone and 50 ng of insert in 5 μL were mixed with 15 μL of Gibson master mix and incubated at $50°$C for 1 hour. The mixture was subsequently transformed with 100 μL of chemical competent *E. coli* and plated in LB agar plates supplemented with 100 μg/mL carbenicillin. Sequencing reactions were carried out to verify the fidelity of the PCR reactions.

6.2.4 DNA precipitation by pellet paint

Per 500 μL of sample, the DNA sample was mixed with 2 μL of pellet paint co-precipitant and 2.5 sample volume of 100% ethanol and vortexed after each addition. After the mixture was incubated on ice for 5 minutes, the DNA was precipitated by centrifugation at maximum speed on a tabletop

Eppendorf centrifuge for 5 minutes at 4°C. The supernatant was aspirated carefully with a pipet, and the pellet was washed with 200 μL of 70% ethanol and vortexed. The sample was again centrifuged at maximum speed for 30 seconds at 4°C, supernatant aspirated, washed with 200 μL of 100% ethanol, and centrifuged at maximum speed for 30 seconds at 4°C. After the supernatant was aspirated to remove the last drop of ethanol, the DNA pellet was dried at 55°C in a tabletop Eppendorf thermomixer, with the cap open, for maximum of 10 minutes. Total removal of ethanol was ensured when the pellet jumped around the tube after light flicking. The DNA pellet was resuspended with sterile deionized water or buffer to the desire concentration.

6.2.5   Chemically competent *E. coli* preparation

One colony of DH5α or XL1-blue was used to inoculate 5 mL of LB medium and grown at 37°C and 250 rpm overnight in orbital shakers. The overnight culture (2 mL) was used to inoculate 200 mL of LB medium and grown for 3 hours at 37°C and 250 rpm or until Abs$_{600}$ reached 0.6 − 0.8. The culture was centrifuged in cold 50 mL conical tubes at 1,200 x g for 5 minutes at 4°C. The supernatant was carefully removed, and the cell pellet was resuspended by pipetting in 8 mL of ice-cold TSS buffer (LB medium at pH 6.5 supplemented with 50 mM magnesium chloride, 10% w/v PEG8000, and 5% v/v DMSO). The cell solution was aliquoted into 100 μL tubes, snap-frozen on dry ice, and stored at -80°C. For transformation, the frozen cells were thawed on ice for about 10 minutes before adding 20 μL of Gibson reaction and incubating on ice for an additional 15 minutes. The cell mixtures were heat shocked at 42°C for 90 seconds on tabletop Eppendorf Thermomixer before returning to ice for 2 minutes. The cell mixtures were rescued by adding 1 mL of warm SOC medium per transformation and incubating at 37°C and 250 rpm for 1 hour. The cells were concentrated to 100 μL by centrifuging at 2,500 x g for 1 minute and removing all but 100 μL of supernatant and plated on LB agar plates supplemented with 100 μg/mL carbenicillin.

6.3 Enzyme purification

6.3.1 Large-scale enzyme production in *S. cerevisiae*

Plasmids containing the desired gene construct were transformed into yeast using Frozen-EZ Yeast Transformation II Kit, and the transformed yeast was grown at 30°C on SD-Ura agar plates for 3 days. Yeast colonies expressing Cel6A with C-terminal His_6 tag or Cel7A with N-terminal His_8 tag were used to inoculate 5 mL to 10 mL SD-Ura medium and grown overnight at 30°C and 250 rpm in an orbital shaker. The overnight culture was expanded in 100 mL to 200 mL YPD medium and grown at 30°C and 250 rpm for an additional 48 hours. Alternatively, the overnight culture was expanded in 100 mL SD-Ura medium and grown at 30°C for 24 hours before expanding in 1 L YPD medium for an additional 48-hour growth at 30°C and 250 rpm.

6.3.2 Large-scale enzyme production in *E. coli*

E. coli BL21 colonies expressing the desired protein construct under *lac* operon were used to inoculate 5 mL of LB medium supplemented with 100 μg/mL of carbenicillin and grown overnight at 37°C and 250 rpm in an orbital shaker. The overnight culture was used to inoculate 300 mL of TB medium supplemented with 4 mL/L glycerol and 100 μg/mL of carbenicillin and grown at 37°C and 250 rpm for 3 hours or until the Abs_{600} reached 1.6 – 1.8. The culture was cooled on ice until the orbital shaker was cooled to 16°C or 20°C. Fresh IPTG stock of 100 mM was made at this point by dissolving 0.12 g of IPTG in 5 mL of sterile water. Once the shaker cooled, protein expression was induced by adding IPTG to a final concentration of 100 μM and grown at 16°C or 20°C and 250 rpm for 24 hours.

6.3.3 Protein harvest

Cellulases expressed in *S. cerevisiae* were harvested by centrifuging the cells at 5,000 x g for 10 minutes at 4°C. The supernatant was collected and filtered with 0.2 μm PES filter unit from Nelgene. If the expected protein yield was low, the pH of the supernatant was adjusted with 10% sodium

hydroxide to pH $\sim$7.

Cellulases expressed in *E. coli* were harvested by centrifuging the cells at 5,000 x g for 10 minutes at 4°C. The cell pellet was resuspended in 5 mL cold lysis buffer per gram of cells (20 mM Tris buffer, pH 8.0, with 100 mM of sodium chloride and 2 mg/L DNAse I) and lysed via sonication on ice. Samples containing 1 μL of unlysed cell suspension were collected for further analysis by SDS-gel electrophoresis. The cell suspension was sonicated at 15 W with 1-minute program, pulsing for 1 s for every 0.5 s. The 1-minute program was repeated several times until the cell solution changed from milky to translucent grey. The soluble fraction was harvested by centrifugation at 27,000 x g for 30 minutes at 4°C and filtered with 0.2 μm PES filter unit from Nelgene. Samples containing 1 μL of soluble fraction were collected for further analysis by SDS-gel electrophoresis.

6.3.4   One-step his-trap purification

Cellulases with terminal His$_6$ tags were purified with HisTrap HP columns, using 20 mM Tris buffer, pH 8.0, supplemented with 100 mM sodium chloride and 10 mM imidazole as binding buffer and 20 mM Tris buffer, pH 8.0, supplemented with 100 mL sodium chloride and 300 mM imidazole as elution buffer. After sample loading, the bound protein was washed with 5 column volume (CV) of binding buffer before eluting with a linear gradient of 0 – 80% elution buffer in 20 CV. The eluted cellulase peak is typically found at 30 – 40% elution buffer ($\sim$100 mM imidazole). The purified cellulase fractions were pooled and buffer-exchanged with 20 mL of 50 mM sodium acetate buffer, pH 5.0, for three times before being concentrated to a final volume of 300 μL to 500 μL.

6.3.5   Desalting and buffer exchange

One method to desalt purified cellulase samples was via Vivaspin 20 device with 30 kDa molecular weight cutoff (MWCO) for full-length Cel6A or Cel7A, or with 10 kDa MWCO for Cel6A catalytic domain. Typically, the Vivaspin column was centrifuged at 4,000 x g for 30 to 60 minutes per 20 mL diluted sample. Alternatively, the pooled fractions from his-trap purification were desalted using HiPrep 26/10 desalting column. The desalting column was first equilibrated with the target buffer

for 1.5 CV before loading ∼2 mL of pooled his-trap fractions onto the column. The volume of sample loaded was adjusted to ensure that the protein was eluted before the salt elution began. The process was repeated until all pooled his-trap fractions were loaded onto the column and buffer-exchanged to the target buffer.

6.3.6 Two-step his-trap anion-exchange purification

Large cultures (1 L and above; purified with 5 mL HisTrap columns) of cellulases often require two-step purification to achieve the same purity as cellulases from smaller cultures (50 – 200 mL; purified with 1 mL HisTrap columns). Cellulases with terminal His_6 tag were purified as described in the one-step his-trap purification method. After his-trap purification, the purified cellulase fractions were pooled and buffer-exchanged with 20 mL of 20 mM Tris buffer, pH 8.0, for three times before concentrated to a final volume of 500 μL. The concentrated protein sample was diluted with 40 mL of 20 mM Tris buffer, pH 8.0, and purified with HiTrap Q HP, using 20 mM Tris buffer, pH 8.0, as binding buffer and 20 mM Tris buffer, pH 8.0, supplemented with 1 M sodium chloride. After sample loading, the bound protein was washed with 5 column volume (CV) of binding buffer before eluting with a linear gradient of 0 – 30% elution buffer in 20 CV. The eluted cellulase peak is typically found at ∼10% elution buffer (∼100 mM sodium chloride). The purified cellulase fractions were pooled and buffer-exchanged with 20 mL of 50 mM sodium acetate buffer, pH 5.0, for three times before concentrated to a final volume of 300 – 500 μL.

6.3.7 Cleavage of N-glycosylation

Samples containing 15 mg of purified cellulases in 50 mM sodium acetate buffer, pH 5.0, were combined with 30 μL of 500 U/μL PNGase F (New England Biolabs) and a final concentration of 50 mM sodium phosphate buffer, pH 7.5 supplied by the manufacture. The reaction was carried out at 37°C for 6 – 8 hours. After the digest, the reaction was diluted with 20 mL of 20 mM Tris buffer, pH 8.0, supplemented with 100 mM of sodium chloride and 10 mM of imidazole and purified with HisTrap columns as described above. After purification, the samples were concentrated to

∼1 mL using the Vivaspin 20 device with 10 kDa MWCO and desalted and buffer-exchanged to 50 mM sodium acetate buffer, pH 5.0, using a PD-10 desalting column. The eluted samples from PD-10 desalting columns (in 3.5 mL in accordance to the protocol from the manufacturer) were concentrated to ∼500 μL. After protein concentration determination, the sample was adjusted to a final concentration of 20 μg/μL, snap-frozen with dry ice, and stored at -80°C.

6.3.8 Protein concentration determination

Protein concentrations were determined by denaturing protein samples in 6 M guanidine hydrochloride, 25 mM sodium phosphate, pH 6.5, at 25°C and measuring the Abs_{280} of ∼5 – 20 μL concentrated protein samples in 1 mL of guanidine hydrochloride. The volume of added protein samples was determined to reach a target Abs_{280} of 0.1. The protein concentrations were calculated using theoretical extinction coefficients determined using the amino acid composition with ProtParam on the ExPASy server [150]. Extinction coefficients for the variants reported in this study can be found in Table 6.1.

6.4 Reducing sugar assays

6.4.1 Nelson-Somogyi reducing sugar assay

For reducing sugar in the range of 0.15 mM to 2 mM, a modified Nelson-Somogyi assay was used [151, 152]. Typically, 50 μL of sugar solution was mixed with 40 μL of the carbonate-tartrate reagent (180 g/L Na_2SO_4, 15 g/L Rochelle salt, 30 g/L Na_2CO_3, and 20 g/L $NaHCO_3$) and 10 μL of the copper reagent (180 g/L Na_2SO_4 and 12.8 g of anhydrous $CuSO_4$) and boiled at 95°C for 15 minutes in a Thermocycler with a heated lid. The reaction was subsequently cooled to 4°C and mixed with 50 μL of the arsenomolybdate solution (50 g/L $(NH_4)_2MoO_4$, 1.5 N H_2SO_4, 6 g/L NaH_2AsO_4, incubated at 37°C for 16 – 24 hours for the formation of the chromogenic compound and stored in the dark). The reagents were mixed thoroughly to ensure the evolution of CO_2 was completed and the maximum color development was achieved. After centrifuging the reagents briefly

Table 6.1: Extinction coefficients used to determine protein concentration.

Enzyme variant	Extinction coefficient (M^{-1} cm^{-1})
Wild-type Cel6A	
HiCel6A	99405
Hi C403S	99405
Hi proline mutants	99405
HjCel6A	98040
Hj C245S C400S	97915
CtCel6A	99405
HJPlus lineage	
HJPlus, 1F4, 1G6, 2B3, 2F4, 2G6, 3C6, 3C6P	92415
317F12	97915
413A1	97915
3C4W	97915
3CW2	103415
Catalytic domain only	
HJPluscat	75080
3C6Pcat	75080
3C6P cysteine mutants	
3C6P C246A, C246G, C246L, C246S	92415
3C6P C246G S400C	92415
CC1	92290
CC2	92290
CC0	92165
3C6P consensus mutants	
3C6P W99Y	88405
3C6P Y186L	90925
3C6P P321W	97915
Rest of the consensus mutants	92415
3C6P proline mutants	
3C6P W364P	86915
Rest of the proline mutants	92415
Cel7A	
TeCel7A	81260
TS8 Cel7A	81260
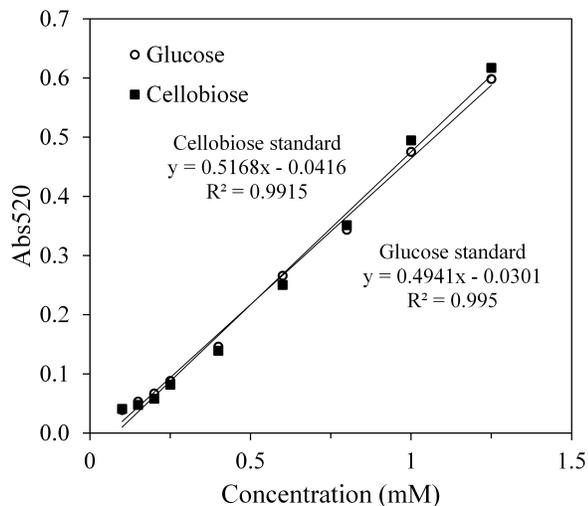
Figure 6.1: Example glucose and cellobiose standards for Nelson-Somogyi reducing sugar assay. Actual absorbance and conversion factor between the sugar concentration and Abs_{520} vary slightly from experiment to experiment.

to remove the CO_2 in the solution, the absorbance of the sugar solution at 520 nm was obtained using a SpectraMax microplate reader with or without cellobiose solution as standard. Example glucose and cellobiose standards and their Abs_{520} can be found in Figure 6.1.

6.4.2 Park-Johnson reducing sugar assay

For reducing sugar in the range of 20 μM to 150 μM, a modified Park-Johnson assay was used [153]. Typically, 50 μL of sugar solution was mixed with 100 μL of the ferricyanide reagent (0.5 g/L $K_3Fe(CN)_6$, 0.2 M K_2HPO_4, pH 10.6) and 50 μL of the carbonate-cyanide reagent (5.3 g/L Na_2CO_3, 0.65 g/L KCN, stored in the dark) and boiled at 95°C for 15 minutes in a thermocycler with a heated lid. The reaction was subsequently cooled to 4°C and mixed with 100 μL of the ferric iron solution (2.5 g/L $FeCl_3$, 10 g/L polyvinylpyrrolidone, 2 N H_2SO_4). After 30 seconds to 5 minutes of color development, the absorbance of the sugar solution at 590 nm was obtained using a SpectraMax microplate reader, with cellobiose solution as standards.

6.4.3 DNS reducing sugar assay

For reducing sugar in the range of 2 mM to 5 mM, a modified DNS assay was used [154]. Typically, 75 μL of sugar solution was mixed with 75 μL of DNS reagent (14 g/L dinitrosalicylic acid, 280 g/L sodium potassium tartrate, 14 g/L NaOH, stored in the dark) and boiled at 95°C for 5 minutes in a Thermocycler with a heat lid. The absorbance of the sugar solution at 540 nm was recorded using a SpectraMax microplate reader.

Appendix A

Wild-type Cel6A and 3C6P Alignment

Multiple sequence alignment between 3C6P and wild-type Cel6A enzyme from *Humicola insolens*, *Hypocrea jecorina*, and *Chaetomium thermophilum* used in this study.

```
H_jecorina       ASCSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPT 60
3C6P             ASCSSVWGQCGGQNWSGPTCCASGSTCVYFNDYYSQCLPGAASSSSSTRAASTTSRVSPT 60
H_insolens       ASCSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPT 60
C_thermophilum   ASCSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPT 60

H_jecorina       TSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTG-A 119
3C6P             TSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFEGVQLWANNYYRSEVHTLAIPQITDPA 120
H_insolens       TSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFEGVQLWANNYYRSEVHTLAIPQITDPA 120
C_thermophilum   TSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFSGVQLWANTYYSSEVHTLAIPSLS-PE 119

H_jecorina       MATAAAAVAKVPSFMWLDT-LDKTPLMEQTLADIRTANKNGGN--YAGQFVVYDLPDRDC 176
3C6P             LRAAASAAAEVPSFLWLDT-LDKTPLMEQTLADIRTANKNGGN--YAGQFVVYDLPDRDC 177
H_insolens       LRAAASAVAEVPSFQWLDRNVTVDTLLVQTLSEIREANQAGANPQYAAQIVVYDLPDRDC 180
C_thermophilum   LAAKAAKVAEVPSFQWLDRNVTVDTLFSGTLAEIRAANQRGANPPYAGIFVVYDLPDRDC 179

H_jecorina       AALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCA 236
3C6P             AALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCA 237
H_insolens       AAAASNGEWAIANNGVNNYKAYINRIREILISFSDVRTILVIEPDSLANMVTNMNVPKCS 240
C_thermophilum   AAAASNGEWSIANNGANNYKRYIDRIRELLIQYSDIRTILVIEPDSLANMVTNMNVQKCS 239

H_jecorina       NAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRAL 296
3C6P             NAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANLDPAAQLFANVYKNASSPRAL 297
H_insolens       GAASTYRELTIYALKQLDLPHVAMYMDAGHAGWLGWPANIQPAAELFAKIYEDAGKPRAV 300
C_thermophilum   NAASTYKELTVYALKQLNLPHVAMYMDAGHAGWLGWPANIQPAAELFAQIYRDAGRPAAV 299

H_jecorina       RGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGK 356
3C6P             RGLATNVANYNAWSIASPPPYTSPNPNYDEKHYIEAFAPLLRNQGFD-AKFIVDTGRNGK 356
H_insolens       RGLATNVANYNAWSVSSPPPYTSPNPNYDEKHYIEAFRPLLEARGFP-AQFIVDQGRSGK 359
C_thermophilum   RGLATNVANYNAWSIASPPSYTSPNPNYDEKHYIEAFAPLLRNQGFD-AKFIVDTGRNGK 358
```

```
H_jecorina        QPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCA 416
3C6P              QPTGQLEWGHWCNVKGTGFGVRPTANTGHELVDAFVWVKPGGESDGTSDPSAPRFDPHCA 416
H_insolens        QPTGQKEWGHWCNAIGTGFGMRPTANTGHQYVDAFVWVKPGGECDGTSDTTAARYDYHCG 419
C_thermophilum    QPTGQLEWGHWCNVKGTGFGVRPTANTGHELVDAFVWVKPGGESDGTSDTSAARYDYHCG 418

H_jecorina        LPDALQPAPQAGAWFQAYFVQLLTNANPSFL 447
3C6P              LPDALQPAPQAGAWFQAYFVQLLTNANPSFL 447
H_insolens        LEDALKPAPEAGQWFNEYFIQLLRNANPPF- 449
C_thermophilum    LSDALTPAPEAGQWFQAYFEQLLINANPP-- 447
```

Appendix B

Example Raw Data for Thermostability Measurements

Example raw data used to determine T_{50} and half-life for Cel6A variants are presented here.
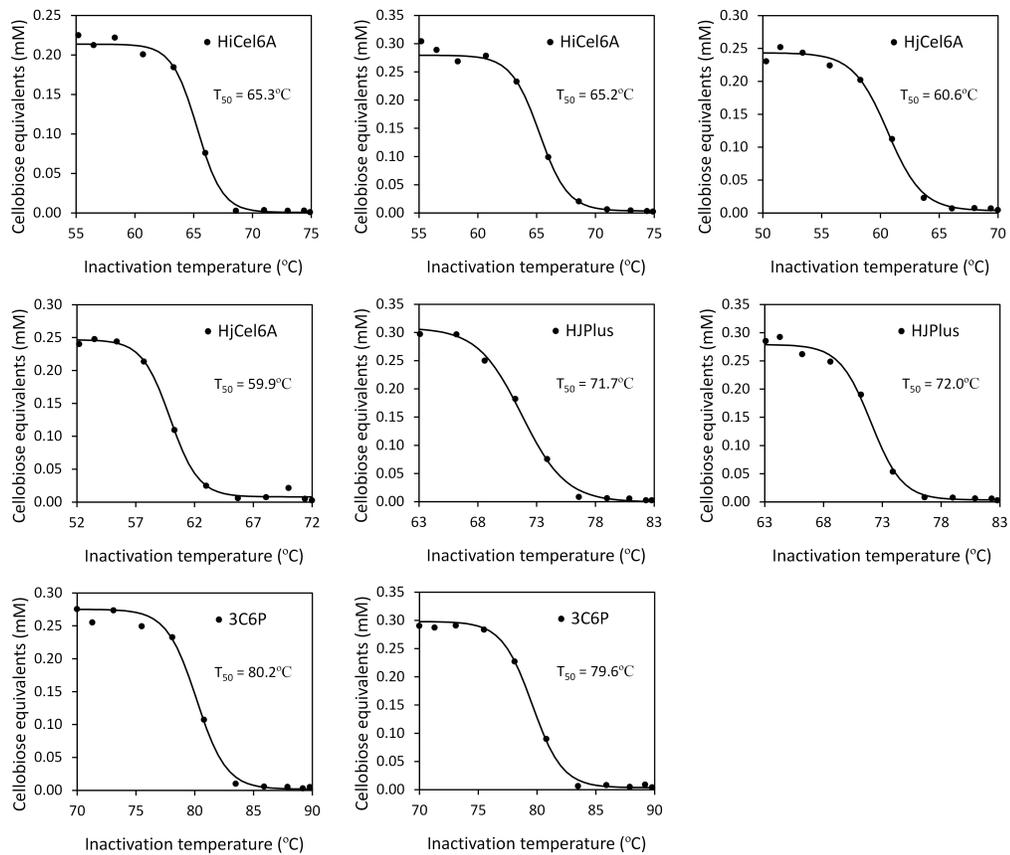

Figure B.1: Example raw data for determining T_{50} of Cel6A. Black dots represent actual data, while the solid lines represent the Boltzmann sigmoidal model.
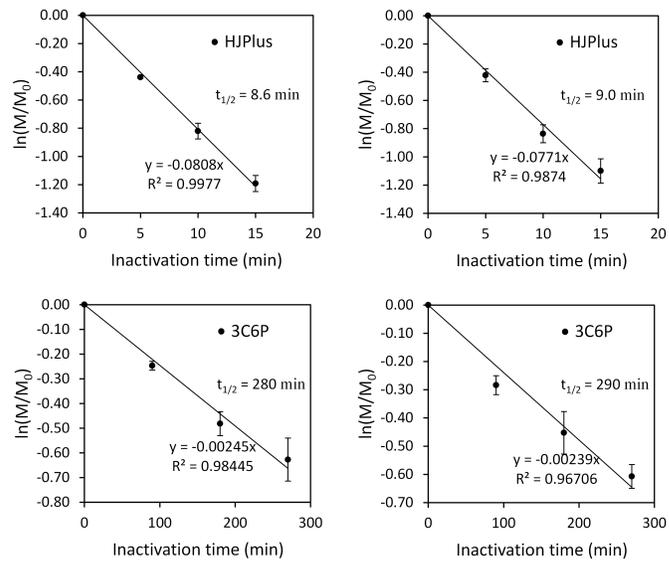
Figure B.2: Example raw data for determining half life, $t_{1/2}$, of Cel6A. Black dots represent actual data, while the solid lines represent the linear model.

Appendix C

Cel6A SCHEMA Recombination Blocks

The Cel6A sequences from *Humicola insolens* (parent 1), *Hypocrea jecorina* (parent 2), and *Chaetomium thermophilum* are listed according to block divisions.

```
Block 1

H_insolens       90 GNPFEGVQLWANNYYRSEVHTLAIPQITDPALRAAASAVAEVPSF 134
H_jecorina       90 GNPFVGVTPWANAYYASEVSSLAIPSLT-GAMATAAAAVAKVPSF 133
C_thermophilum   90 GNPFSGVQLWANTYYSSEVHTLAIPSLS-PELAAKAAKVAEVPSF 133

Block 2

H_insolens      135 QWLDRNVTVDTLLVQTLSEIREANQAG 161
H_jecorina      134 MWLD-TLDKTPLMEQTLADIRTANKNG 159
C_thermophilum  134 QWLDRNVTVDTLFSGTLAEIRAANQRG 160

Block 3

H_insolens      162 ANPQYAAQIVVYDLPDRDCAAAASNGEWAIANNGVNNYKAYIN 204
H_jecorina      160 GN--YAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYID 200
C_thermophilum  161 ANPPYAGIFVVYDLPDRDCAAAASNGEWSIANNGANNYKRYID 203

Block 4

H_insolens      205 RIREILISFSDVRTILVIEPDSLANMVTNMNVPKCSGAAS 244
H_jecorina      201 TIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQS 240
C_thermophilum  204 RIRELLIQYSDIRTILVIEPDSLANMVTNMNVQKCSNAAS 243
```

Block 5

```
H_insolens      245 TYRELTIYALKQLDLPHVAMYMDAGHAGWLGWPANIQPAAELFAKIYEDAGKPRAVR 301
H_jecorina      241 AYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRALR 297
C_thermophilum  244 TYKELTVYALKQLNLPHVAMYMDAGHAGWLGWPANIQPAAELFAQIYRDAGRPAAVR 300
```

```
H_insolens      302 GLATNVANYN 311
H_jecorina      298 GLATNVANYN 307
C_thermophilum  301 GLATNVANYN 310
```

Block 6

```
H_insolens      312 AWSVSSPPPYTSPNPNYDEKHYIEAFRPLLEARGFP 347
H_jecorina      308 GWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWS 343
C_thermophilum  311 AWSIASPPSYTSPNPNYDEKHYIEAFAPLLRNQGFD 346
```

Block 7

```
H_insolens      348 -AQFIVDQGRSGKQPTGQKEWGHWCNAIGTGFGMRPTANTGHQYVDAFVWVKPGGEC 403
H_jecorina      344 NAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGEC 400
C_thermophilum  347 -AKFIVDTGRNGKQPTGQLEWGHWCNVKGTGFGVRPTANTGHELVDAFVWVKPGGES 402
```

Block 8

```
H_insolens      404 DGTSDTTAARYDYHCGLEDALKPAPEAGQWFNEYFIQLLRNANPPF- 449
H_jecorina      401 DGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL 447
C_thermophilum  403 DGTSDTSAARYDYHCGLSDALTPAPEAGQWFQAYFEQLLINANPP-- 447
```

Appendix D

Cel6 Multiple Sequence Alignment

D.1 Consensus sequence

The alignment between 3C6P and the consensus sequence from the multiple sequence alignment is presented here. The consensus sequence is composed of amino acids that are present in more than 50% of the aligned sequences or '-' if no consensus is found at each position.

```
3C6P        GNPFEGVQLWANNYYRSEVHTLAIPQITDPALRAAASAAAEVPSFLWLDTLDKTPLMEQT 60
Consensus    GNPF-G---Y-NP-Y--EV---------D--LAAKA--VA--PTFVWLD-------L--- 27
            **** *   : *   *   **         *   *  *  *   .*   *:*:***      :

3C6P        TANKNGGNYAG--QFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIV 118
Consensus    L--A------G--PP---QFV-YDLP-RDCAALASNGEL-----G---YK-YID-I---L 59
            *         *   .   *** **** ***********     *   ** *** *   :

3C6P        VEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGH 178
Consensus    --Y-D-R---VIEPDSL-NLVTN-----C--A---Y--G--YAL--L--PNV--Y-DAGH 90
             * * *    ******* *****     *   *    *       **:   *   ***   * ****

3C6P        AGWLGWPANLDPAAQLFANVYKNASSPRALRGLATNVANYNAWSIASPPPYTSPNPNYDE 238
Consensus    -GWLGW--N--P-A-LF------AG----VRG--TNVANY-----------T-WN---DE 116
             *****   *   * * **       *.      :**   ******            *   *    **

3C6P        KHYIEAFAPLLRNQGFDAKFIVDTGRNGKQPTGQLEWGHWCNVKGTGFGVRPTANTGHEL 298
Consensus    --Y--AL---L---G----FI-DTGRNG--G-RR---G-WCN--GAG-G-RP-A-TG--- 145
              *  *:   *   *     ** ******       :    * ***   *:* * ** * **

3C6P        VDAFVWVKPGGESDGTSDPSAPRFDPHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL 358
Consensus    -DA-VWVKP-GESDG-S-----R-D--C----AL--AP-AG-WF-AYF--L--NA-P--L 177
             ** ***** ***** *       * * *      **  ** ** ** ***   * ** *   *
```

D.2   Genbank ID

The GenBank IDs for the Cel6 sequences used in the multiple sequence alignment are listed here. Cel6A sequences from *H. insolens*, *H. jecorina*, and *C. thermophilum* were used as the query sequence for Protein BLAST.

Table D.1: List of GenBank IDs for the Cel6 sequences used in the multiple sequence alignment.

H_insolens	H_jecorina	C_thermophilum	XP_001903645.1	ABY52793.1
AAM76664.1	ZP_06907851.1	ACI26721.1	XP_001258843.1	ZP_06710849.1
BAF80327.1	YP_003342336	AAL01212.1	XP_001841132	ZP_07297522.1
ZP_16176803.1	XP_001598803.1	XP_748511.1	XP_001795501.1	XP_002484367.1
AAU05379.2	YP_003555625.1	ZP_07269959.1	AAD41097.1	XP_662886.1
ABY52799.1	ZP_06922080.1	AAD02028.1	YP_431432.1	YP_004453442.1
ZP_07309615.1	YP_003487397	YP_003511715	YP_003325739.1	AAC09067.1
YP_003494467.1	NP_823029.1	XP_001934153.1	XP_001226518.1	XP_001264772.1
YP_003636992.1	AAB92679.1	XP_001392295.2	AAQ76094.1	ABS72374.1
ACH96126.1	P49075.1	YP_003653250.1	1OCN_A	ABG48766.1
AAB92678.1	XP_002391276.1	AAK28357.1	AAL92497.1	AAF34679.1
P46236.1	YP_001912004.1	XP_001210279.1	ZP_06710849	AAF35251.1
ZP_06485553.1	XP_001933777.1	Q9C1S9.1	AAW64927.1	YP_001544904.1
YP_001828440.1	YP_003342336.1	ZP_07284935.1	ZP_06920986.1	ABY52798.1
ZP_12210081.1	ZP_06920986	BAH08705.1	1HGY_A	Q7SIG5.1
NP_630629.1	YP_003099982.1	ABY52797.1	BAH59082.1	1OC5_A
ACZ34301.1	YP_001544904	YP_003074285.1	Q5B2E8.2	YP_872374.1
XP_001806560.1	XP_003000565.1	ZP_08240659	XP_001216114.1	NP_298556.1
XP_001269265.1	NP_522144.1	YP_003117370.1	AAQ72468.1	XP_001226029.1

Table D.1 – *Continued from previous page*

YP_004582927.1	XP_001552807.1	YP_003115582.1	XP_001836853	YP_003099982
YP_003555620.1	ZP_08240659.1	XP_960770.1	XP_002149891.1	XP_383804.1
YP_004926939.1	AAG39980.1	YP_003375006.1	AAP33843.1	BAL63102.1
AAP30749.1	XP_002380541.1	AAL01211.1	NP_638506.1	XP_001839612
AAK95564.1	YP_001775232.1	XP_001903209.1	ZP_04605447.1	XP_002999918.1
XP_001836853.1	XP_002560902.1	XP_001796781.1	AAD51054.2	AAQ09256.1
YP_003160702.1	AAT64008.1	ZP_06575512.1	YP_004582311.1	1HGW_A
YP_003075283.1	XP_003710956.1	BAA74458.1	XP_003344888.1	XP_001841132.1
XP_001903170.1	AAC49315.1	YP_001982934.1	ZP_00683852.1	AAL15038.1
XP_003344598.1	ADC83999.1	YP_003075283	AAQ09258.1	XP_001931623.1
YP_452834.1	XP_001839612.1	YP_003487397.1	AAC09066	YP_003494467
YP_003838150.1	YP_001538764.1	YP_003074285	XP_001903893.1	XP_003710385.1
XP_003346794.1	4B4F_A	AAA72922.1	YP_003511715.1	YP_001160411.1
1CB2_A	XP_003006976.1	BAH59083.1	NP_638880.1	JC7931
ZP_06488345.1	AAC09066.1	ZP_16328686.1	3VOG_A	ABY52799
AAD51055.1	AAO47726.1	XP_956581.1	XP_957415.1	ZP_00944606.1
ZP_06922080	YP_004800724.1	XP_003049522.1	XP_001395308.1	YP_001828440
ZP_06827307.1	ZP_07297522	YP_003680375.1	BAB83928	AAQ93324.1
BAG48183.1	ADO33720.1	XP_003004556.1	ZP_16176803	BAB83928.1
AAR08200.1	ZP_07307649.1	XP_001273717.1	NP_821732.1	XP_002392869.1
ACH91035.1	XP_001226566.1	AAA50608.1	XP_001792324.1	YP_527744.1
CBF87814.1	XP_003713150.1	YP_001618727.1	XP_001224490.1	XP_003713150.1
YP_001618727.1	XP_001224490.1	ZP_01462143.1		

Appendix E

List of Abbreviations

Words and phrases abbreviated in this thesis can be found in the following.

Table E.1: List of abbreviations.

BSA	Bovine serum albumin
CBM	Cellulose binding module
CD	Circular dichroism
Cel6A	Cellobiohydrolase, family 6
Cel7A	Cellobiohydrolase, family 7
CtCel6A	Cel6A from *Chaetomium thermophilum*
CV	Column volume
Da or kDa	Dalton or kilo-dalton
DNS	Dinitrosalicylic acid reducing sugar assay
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EMIM	1-ethyl-3-methylimidazolium
GH	Glycoside hydrolase
HiCel6A	Cel6A from *Humicola insolens*
HjCel6A	Cel6A from *Hypocrea jecorina*
IC_{50}	Inhibitor concentration with 50% activity
IL	Ionic liquid
ILSG	IL-pretreated switchgrass
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth
MWCO	Molecular weight cutoff
rpm	Rotation per minute
SD-Ura	Synthetic defined medium without uracil
SD-CAA	Synthetic defined medium with casamino acids
SOC	Super optimal broth
$t_{1/2}$	Half-life
T_{50}	Temperature with 50% residual activity
T_{A50}	Temperature with half-maximal activity
TeCel7A	Cel7A from *Talaromyces emersonii*
T_m	Melting temperature
YPD	Yeast extract, peptone, dextrose medium

Bibliography

[1] US Energy Information Adminstration, "Monthly energy review," Feb 2013.

[2] US Energy Information Adminstration, "Annual energy outlook early release overview," 2013.

[3] D. G. Wilson, "Energy supplies and future engines for land, sea, and air.," *J Air Waste Manag Assoc*, vol. 62, pp. 607–624, Jun 2012.

[4] S. Chu and A. Majumdar, "Opportunities and challenges for a sustainable energy future.," *Nature*, vol. 488, pp. 294–303, Aug 2012.

[5] R. J. Cogdell, A. T. Gardiner, and L. Cronin, "Learning from photosynthesis: how to use solar energy to make fuels.," *Philos Transact A Math Phys Eng Sci*, vol. 370, pp. 3819–3826, Aug 2012.

[6] R. E. H. Sims, W. Mabee, J. N. Saddler, and M. Taylor, "An overview of second generation biofuel technologies.," *Bioresour Technol*, vol. 101, pp. 1570–1580, Mar 2010.

[7] Y.-H. P. Zhang and L. R. Lynd, "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems.," *Biotechnol Bioeng*, vol. 88, pp. 797–824, Dec 2004.

[8] H. A. Krassig, *Cellulose: structure, accessibility and reactivity.* Yverdon, Switzerland:: Gordon & Breach, 1993.

[9] Y.-H. P. Zhang, M. E. Himmel, and J. R. Mielenz, "Outlook for cellulase improvement: screening and selection strategies.," *Biotechnol Adv*, vol. 24, no. 5, pp. 452–481, 2006.

[10] R. Wolfenden, X. Lu, and G. Young, "Spontaneous hydrolysis of glycosides," *J Am Chem Soc*, vol. 120, no. 27, pp. 6814–6815, 1998.

[11] H. W. Blanch, B. A. Simmons, and D. Klein-Marcuschamer, "Biomass deconstruction to sugars.," *Biotechnol J*, vol. 6, pp. 1086–1102, Sep 2011.

[12] V. S. Chang and M. T. Holtzapple, "Fundamental factors affecting biomass enzymatic reactivity.," *Appl Biochem Biotechnol*, vol. 84-86, pp. 5–37, 2000.

[13] L. da Costa Sousa, S. P. S. Chundawat, V. Balan, and B. E. Dale, "'cradle-to-grave' assessment of existing lignocellulose pretreatment technologies.," *Curr Opin Biotechnol*, vol. 20, pp. 339–347, Jun 2009.

[14] J. Y. Zhu, X. Pan, and R. S. Zalesny, Jr, "Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance.," *Appl Microbiol Biotechnol*, vol. 87, pp. 847–857, Jul 2010.

[15] L. Tao, A. Aden, R. T. Elander, V. R. Pallapolu, Y. Y. Lee, R. J. Garlock, V. Balan, B. E. Dale, Y. Kim, N. S. Mosier, M. R. Ladisch, M. Falls, M. T. Holtzapple, R. Sierra, J. Shi, M. A. Ebrik, T. Redmond, B. Yang, C. E. Wyman, B. Hames, S. Thomas, and R. E. Warner, "Process and technoeconomic analysis of leading pretreatment technologies for lignocellulosic ethanol production using switchgrass.," *Bioresour Technol*, vol. 102, pp. 11105–11114, Dec 2011.

[16] T. A. Lloyd and C. E. Wyman, "Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids.," *Bioresour Technol*, vol. 96, pp. 1967–1977, Dec 2005.

[17] S. C. Rabelo, R. Maciel Filho, and A. C. Costa, "Lime pretreatment and fermentation of enzymatically hydrolyzed sugarcane bagasse.," *Appl Biochem Biotechnol*, Jan 2013.

[18] H. Wang, G. Gurau, and R. D. Rogers, "Ionic liquid processing of cellulose.," *Chem Soc Rev*, vol. 41, pp. 1519–1537, Feb 2012.

[19] P. Tomme, R. A. Warren, and N. R. Gilkes, "Cellulose hydrolysis by bacteria and fungi.," *Adv Microb Physiol*, vol. 37, pp. 1–81, 1995.

[20] C. Divne, J. Sthlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. K. Knowles, T. T. Teeri, and T. A. Jones, "The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*.," *Science*, vol. 265, pp. 524–528, Jul 1994.

[21] J. Rouvinen, T. Bergfors, T. Teeri, J. K. Knowles, and T. A. Jones, "Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*.," *Science*, vol. 249, pp. 380–386, Jul 1990.

[22] C. Boisset, C. Fraschini, M. Schulein, B. Henrissat, and H. Chanzy, "Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A.," *Appl Environ Microbiol*, vol. 66, pp. 1444–1452, Apr 2000.

[23] J. Kraulis, G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, and A. M. Gronenborn, "Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing.," *Biochemistry*, vol. 28, pp. 7241–7257, Sep 1989.

[24] J. Lehtio, J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder, and T. T. Teeri, "The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules.," *Proc Natl Acad Sci U S A*, vol. 100, pp. 484–489, Jan 2003.

[25] M. Couturier, J. Feliu, M. Haon, D. Navarro, L. Lesage-Meessen, P. M. Coutinho, and J.-G. Berrin, "A thermostable GH45 endoglucanase from yeast: impact of its atypical multimodularity on activity.," *Microb Cell Fact*, vol. 10, p. 103, 2011.

[26] M. R. Nimlos, G. T. Beckham, J. F. Matthews, L. Bu, M. E. Himmel, and M. F. Crowley, "Binding preferences, surface attachment, diffusivity, and orientation of a family 1

carbohydrate-binding module on cellulose.," *J Biol Chem*, vol. 287, pp. 20603–20612, Jun 2012.

[27] G. T. Beckham, J. F. Matthews, Y. J. Bomble, L. Bu, W. S. Adney, M. E. Himmel, M. R. Nimlos, and M. F. Crowley, "Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase.," *J Phys Chem B*, vol. 114, pp. 1447–1453, Jan 2010.

[28] L. J. Gibson, "The hierarchical structure and mechanics of plant materials.," *J R Soc Interface*, vol. 9, pp. 2749–2766, Nov 2012.

[29] Z. Wang, T. Lan, and J. Zhu, "Lignosulfonate and elevated pH can enhance enzymatic saccharification of lignocelluloses.," *Biotechnol Biofuels*, vol. 6, no. 1, p. 9, 2013.

[30] A. Varrot, J. Macdonald, R. V. Stick, G. Pell, H. J. Gilbert, and G. J. Davies, "Distortion of a cellobio-derived isofagomine highlights the potential conformational itinerary of inverting β-glucosidases.," *Chem Commun (Camb)*, pp. 946–947, Apr 2003.

[31] G. J. Davies, A. M. Brzozowski, M. Dauter, A. Varrot, and M. Schulein, "Structure and function of *Humicola insolens* family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 å resolution.," *Biochem J*, vol. 348 Pt 1, pp. 201–207, May 2000.

[32] A. Varrot, M. Schulein, and G. J. Davies, "Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding.," *Biochemistry*, vol. 38, pp. 8884–8891, Jul 1999.

[33] A. Varrot, S. Hastrup, M. Schulein, and G. J. Davies, "Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 å resolution.," *Biochem J*, vol. 337 (Pt 2), pp. 297–304, Jan 1999.

[34] J. Medve, J. Stahlberg, and F. Tjerneld, "Adsorption and synergism of cellobiohydrolase I and II of *Trichoderma reesei* during hydrolysis of microcrystalline cellulose.," *Biotechnol Bioeng*, vol. 44, pp. 1064–1073, Nov 1994.

[35] M. Schulein, "Enzymatic properties of cellulases from *Humicola insolens.*," *J Biotechnol*, vol. 57, pp. 71–81, Sep 1997.

[36] A. Koivula, T. Reinikainen, L. Ruohonen, A. Valkeajarvi, M. Claeyssens, O. Teleman, G. J. Kleywegt, M. Szardenings, J. Rouvinen, T. A. Jones, and T. T. Teeri, "The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169.," *Protein Eng*, vol. 9, pp. 691–699, Aug 1996.

[37] A. Varrot, T. P. Frandsen, I. von Ossowski, V. Boyer, S. Cottaz, H. Driguez, M. Schulein, and G. J. Davies, "Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens.*," *Structure*, vol. 11, pp. 855–864, Jul 2003.

[38] J. Jalak and P. Valjamae, "Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis.," *Biotechnol Bioeng*, vol. 106, pp. 871–883, Aug 2010.

[39] B. Yang, D. M. Willies, and C. E. Wyman, "Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion.," *Biotechnol Bioeng*, vol. 94, pp. 1122–1128, Aug 2006.

[40] K. Igarashi, T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttila, T. Ando, and M. Samejima, "Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface.," *Science*, vol. 333, pp. 1279–1282, Sep 2011.

[41] R. Maheshwari, G. Bharadwaj, and M. K. Bhat, "Thermophilic fungi: their physiology and enzymes.," *Microbiol Mol Biol Rev*, vol. 64, pp. 461–488, Sep 2000.

[42] A. J. Shaw, K. K. Podkaminer, S. G. Desai, J. S. Bardsley, S. R. Rogers, P. G. Thorne, D. A. Hogsett, and L. R. Lynd, "Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield.," *Proc Natl Acad Sci U S A*, vol. 105, pp. 13769–13774, Sep 2008.

[43] B. M. A. Abdel-Banat, H. Hoshida, A. Ano, S. Nonklang, and R. Akada, "High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast?," *Appl Microbiol Biotechnol*, vol. 85, pp. 861–867, Jan 2010.

[44] P. Turner, G. Mamo, and E. N. Karlsson, "Potential and utilization of thermophiles and thermostable enzymes in biorefining.," *Microb Cell Fact*, vol. 6, p. 9, 2007.

[45] P. Heinzelman, C. D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull, and F. H. Arnold, "A family of thermostable fungal cellulases created by structure-guided recombination.," *Proc Natl Acad Sci U S A*, vol. 106, pp. 5610–5615, Apr 2009.

[46] J. B. Endelman, J. J. Silberg, Z.-G. Wang, and F. H. Arnold, "Site-directed protein recombination as a shortest-path problem.," *Protein Eng Des Sel*, vol. 17, pp. 589–594, Jul 2004.

[47] R. C. Kuhad, R. Gupta, and A. Singh, "Microbial cellulases and their industrial applications.," *Enzyme Res*, vol. 2011, p. 280696, 2011.

[48] D. B. Jordan, M. J. Bowman, J. D. Braker, B. S. Dien, R. E. Hector, C. C. Lee, J. A. Mertens, and K. Wagschal, "Plant cell walls to ethanol.," *Biochem J*, vol. 442, pp. 241–252, Mar 2012.

[49] D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, and H. W. Blanch, "The challenge of enzyme cost in the production of lignocellulosic biofuels.," *Biotechnol Bioeng*, vol. 109, pp. 1083–1087, Apr 2012.

[50] C. Fu, J. R. Mielenz, X. Xiao, Y. Ge, C. Y. Hamilton, M. Rodriguez, Jr, F. Chen, M. Foston, A. Ragauskas, J. Bouton, R. A. Dixon, and Z.-Y. Wang, "Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass.," *Proc Natl Acad Sci U S A*, vol. 108, pp. 3803–3808, Mar 2011.

[51] J. H. Jung, W. M. Fouad, W. Vermerris, M. Gallo, and F. Altpeter, "RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass.," *Plant Biotechnol J*, Aug 2012.

[52] Y. Chen, M. A. Stevens, Y. Zhu, J. Holmes, G. Moxley, and H. Xu, "Reducing acid in dilute acid pretreatment and the impact on enzymatic saccharification.," *J Ind Microbiol Biotechnol*, vol. 39, pp. 691–700, May 2012.

[53] J. R. Cherry and A. L. Fidantsef, "Directed evolution of industrial enzymes: an update.," *Curr Opin Biotechnol*, vol. 14, pp. 438–443, Aug 2003.

[54] M. Anbar and E. A. Bayer, "Approaches for improving thermostability characteristics in cellulases.," *Methods Enzymol*, vol. 510, pp. 261–271, 2012.

[55] L. Viikari, M. Alapuranen, T. Puranen, J. Vehmaanpera, and M. Siika-Aho, "Thermostable enzymes in lignocellulose hydrolysis.," *Adv Biochem Eng Biotechnol*, vol. 108, pp. 121–145, 2007.

[56] S. Ferdjani, M. Ionita, B. Roy, M. Dion, Z. Djeghaba, C. Rabiller, and C. Tellier, "Correlation between thermostability and stability of glycosidases in ionic liquid.," *Biotechnol Lett*, vol. 33, pp. 1215–1219, Jun 2011.

[57] S. J. Horn, G. Vaaje-Kolstad, B. Westereng, and V. G. Eijsink, "Novel enzymes for the degradation of cellulose.," *Biotechnol Biofuels*, vol. 5, no. 1, p. 45, 2012.

[58] J. A. Langston, T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko, and M. D. Sweeney, "Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61.," *Appl Environ Microbiol*, vol. 77, pp. 7007–7015, Oct 2011.

[59] R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J.-C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jorgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, and P. H. Walton, "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components.," *Proc Natl Acad Sci U S A*, vol. 108, pp. 15079–15084, Sep 2011.

[60] M. Saloheimo, M. Paloheimo, S. Hakola, J. Pere, B. Swanson, E. Nyyssnen, A. Bhatia, M. Ward, and M. Penttila, "Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials.," *Eur J Biochem*, vol. 269, pp. 4202–4211, Sep 2002.

[61] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.," *Nucleic Acids Res*, vol. 37, pp. D233–D238, Jan 2009.

[62] P. Heinzelman, C. D. Snow, M. A. Smith, X. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull, and F. H. Arnold, "SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability.," *J Biol Chem*, vol. 284, pp. 26229–26233, Sep 2009.

[63] S. E. Lantz, F. Goedegebuur, R. Hommes, T. Kaper, B. R. Kelemen, C. Mitchinson, L. Wallace, J. Stahlberg, and E. A. Larenas, "*Hypocrea jecorina* CEL6A protein engineering.," *Biotechnol Biofuels*, vol. 3, p. 20, 2010.

[64] J. Woodward, M. K. Hayes, and N. E. Lee, "Hydrolysis of cellulose by saturating and non-saturating concentrations of cellulase: Implications for synergism," *Nat Biotechnol*, vol. 6, pp. 301–304, 1988.

[65] L. Rosgaard, S. Pedersen, J. R. Cherry, P. Harris, and A. S. Meyer, "Efficiency of new fungal cellulase systems in boosting enzymatic degradation of barley straw lignocellulose.," *Biotechnol Prog*, vol. 22, no. 2, pp. 493–498, 2006.

[66] S. Suwannarangsee, B. Bunterngsook, J. Arnthong, A. Paemanee, A. Thamchaipenet, L. Eurwilaichitr, N. Laosiripojana, and V. Champreda, "Optimisation of synergistic biomass-degrading enzyme systems for efficient rice straw hydrolysis using an experimental mixture design.," *Bioresour Technol*, vol. 119, pp. 252–261, Sep 2012.

[67] R. S. Komor, P. A. Romero, C. B. Xie, and F. H. Arnold, "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods.," *Protein Eng Des Sel*, vol. 25, pp. 827–833, Dec 2012.

[68] R. Georgescu, G. Bandara, and L. Sun, "Saturation mutagenesis.," *Methods Mol Biol*, vol. 231, pp. 75–83, 2003.

[69] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup, "Isolating and engineering human antibodies using yeast surface display.," *Nat Protoc*, vol. 1, no. 2, pp. 755–768, 2006.

[70] W. Kabsch, "XDS.," *Acta Crystallogr D Biol Crystallogr*, vol. 66, pp. 125–132, Feb 2010.

[71] P. Evans, "Scaling and assessment of data quality.," *Acta Crystallogr D Biol Crystallogr*, vol. 62, pp. 72–82, Jan 2006.

[72] A. Vagin and A. Teplyakov, "Molecular replacement with MOLREP.," *Acta Crystallogr D Biol Crystallogr*, vol. 66, pp. 22–25, Jan 2010.

[73] Collaborative Computational Project, Number 4, "The CCP4 suite: programs for protein crystallography.," *Acta Crystallogr D Biol Crystallogr*, vol. 50, pp. 760–763, Sep 1994.

[74] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics.," *Acta Crystallogr D Biol Crystallogr*, vol. 60, pp. 2126–2132, Dec 2004.

[75] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method.," *Acta Crystallogr D Biol Crystallogr*, vol. 53, pp. 240–255, May 1997.

[76] G. Carrard and M. Linder, "Widely different off rates of two closely related cellulose-binding domains from *Trichoderma reesei*.," *Eur J Biochem*, vol. 262, pp. 637–643, Jun 1999.

[77] P. Heinzelman, R. Komor, A. Kannan, P. Romero, X. Yu, S. Mohler, C. Snow, and F. H. Arnold, "Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination," *Protein Eng Des Sel*, vol. 23, no. 11, pp. 871–80, 2010.

[78] M. Saloheimo and T. M. Pakula, "The cargo and the transport system: secreted proteins and protein secretion in *Trichoderma reesei* (*Hypocrea jecorina*).," *Microbiology*, vol. 158, pp. 46–57, Jan 2012.

[79] N. J. Parry, D. E. Beever, E. Owen, W. Nerinckx, M. Claeyssens, J. Van Beeumen, and M. K. Bhat, "Biochemical characterization and mode of action of a thermostable endoglucanase purified from *Thermoascus aurantiacus*.," *Arch Biochem Biophys*, vol. 404, pp. 243–253, Aug 2002.

[80] C. M. Dana, P. Saija, S. M. Kal, M. B. Bryan, H. W. Blanch, and D. S. Clark, "Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A.," *Biotechnol Bioeng*, vol. 109, pp. 2710–2719, Nov 2012.

[81] C. Boisset, C. Petrequin, H. Chanzy, B. Henrissat, and M. Schulein, "Optimized mixtures of recombinant *Humicola insolens* cellulases for the biodegradation of crystalline cellulose.," *Biotechnol Bioeng*, vol. 72, pp. 339–345, Feb 2001.

[82] C. Dumon, A. Varvak, M. A. Wall, J. E. Flint, R. J. Lewis, J. H. Lakey, C. Morland, P. Luginbuhl, S. Healey, T. Todaro, G. DeSantis, M. Sun, L. Parra-Gessert, X. Tan, D. P. Weiner, and H. J. Gilbert, "Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure.," *J Biol Chem*, vol. 283, pp. 22557–22564, Aug 2008.

[83] M. J. Cuneo, Y. Tian, M. Allert, and H. W. Hellinga, "The backbone structure of the thermophilic *Thermoanaerobacter tengcongensis* ribose binding protein is essentially identical to its mesophilic *E. coli* homolog.," *BMC Struct Biol*, vol. 8, p. 20, 2008.

[84] D. B. Wilson, "Cellulases and biofuels.," *Curr Opin Biotechnol*, vol. 20, pp. 295–299, Jun 2009.

[85] B. Seiboth, C. Ivanova, and V. Seidl-Seiboth, *Biofuel Production-Recent Developments and Prospects*, ch. 13: Trichoderma reesei: A Fungal Enzyme Producer for Cellulosic Biofuels. InTech, 2011.

[86] M. Anbar, O. Gul, R. Lamed, U. O. Sezerman, and E. A. Bayer, "Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis.," *Appl Environ Microbiol*, vol. 78, pp. 3458–3464, May 2012.

[87] Y. Ito, A. Ikeuchi, and C. Imamura, "Advanced evolutionary molecular engineering to produce thermostable cellulase by using a small but efficient library.," *Protein Eng Des Sel*, vol. 26, pp. 73–79, Jan 2013.

[88] I. Wu and F. H. Arnold, "Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures," *Biotechnol Bioeng*, 2013.

[89] C. Vieille and G. J. Zeikus, "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability.," *Microbiol Mol Biol Rev*, vol. 65, pp. 1–43, Mar 2001.

[90] S. E. Blumer-Schuette, I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly, "Extremely thermophilic microorganisms for biomass conversion: status and prospects.," *Curr Opin Biotechnol*, vol. 19, pp. 210–217, Jun 2008.

[91] S. E. Zale and A. M. Klibanov, "Why does ribonuclease irreversibly inactivate at high temperatures?," *Biochemistry*, vol. 25, pp. 5432–5444, Sep 1986.

[92] S. M. Marino and V. N. Gladyshev, "Analysis and functional prediction of reactive cysteine residues.," *J Biol Chem*, vol. 287, pp. 4419–4425, Feb 2012.

[93] M. V. Trivedi, J. S. Laurence, and T. J. Siahaan, "The role of thiols and disulfides on protein stability.," *Curr Protein Pept Sci*, vol. 10, pp. 614–625, Dec 2009.

[94] H. R. Costantino, R. Langer, and A. M. Klibanov, "Solid-phase aggregation of proteins under pharmaceutically relevant conditions.," *J Pharm Sci*, vol. 83, pp. 1662–1669, Dec 1994.

[95] J. R. Lepock, H. E. Frey, and R. A. Hallewell, "Contribution of conformational stability and reversibility of unfolding to the increased thermostability of human and bovine superoxide dismutase mutated at free cysteines.," *J Biol Chem*, vol. 265, pp. 21612–21618, Dec 1990.

[96] M. Sandgren, P. J. Gualfetti, C. Paech, S. Paech, A. Shaw, L. S. Gross, M. Saldajeno, G. I. Berglund, T. A. Jones, and C. Mitchinson, "The *Humicola grisea* Cel12A enzyme structure at 1.2 å resolution and the impact of its free cysteine residues on thermal stability," *Protein Sci*, vol. 12, pp. 2782–2793, Dec 2003.

[97] T. Arakawa, S. J. Prestrelski, W. C. Kenney, and J. F. Carpenter, "Factors affecting short-term and long-term stabilities of proteins.," *Adv Drug Deliv Rev*, vol. 46, pp. 307–326, Mar 2001.

[98] S. Zhang, B. K. Barr, and D. B. Wilson, "Effects of noncatalytic residue mutations on substrate specificity and ligand binding of *Thermobifida fusca* endocellulase Cel6A.," *Eur J Biochem*, vol. 267, pp. 244–252, Jan 2000.

[99] A. M. Larsson, T. Bergfors, E. Dultz, D. C. Irwin, A. Roos, H. Driguez, D. B. Wilson, and T. A. Jones, "Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: the role of tyrosine Y73 in substrate ring distortion.," *Biochemistry*, vol. 44, pp. 12915–12922, Oct 2005.

[100] Y. Liu, M. Yoshida, Y. Kurakata, T. Miyazaki, K. Igarashi, M. Samejima, K. Fukuda, A. Nishikawa, and T. Tonozuka, "Crystal structure of a glycoside hydrolase family 6 enzyme, CcCel6C, a cellulase constitutively produced by *Coprinopsis cinerea.*," *FEBS J*, vol. 277, pp. 1532–1542, Mar 2010.

[101] A. Meinke, H. G. Damude, P. Tomme, E. Kwan, D. G. Kilburn, J. Robert C. Miller, R. A. J. Warren, and N. R. Gilkes, "Enhancement of the endo-β-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop," *J Biol Chem*, vol. 270, no. 9, pp. 4383–6, 1995.

[102] A. J. Miles, F. Wien, J. G. Lees, A. Rodger, R. W. Janes, and B. Wallace, "Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers," *Spectroscopy*, vol. 17, pp. 653–661, 2003.

[103] S. Capasso and S. Salvadori, "Effect of the three-dimensional structure on the deamidation reaction of ribonuclease A.," *J Pept Res*, vol. 54, pp. 377–382, Nov 1999.

[104] H. Boer, T. T. Teeri, and A. Koivula, "Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters.," *Biotechnol Bioeng*, vol. 69, pp. 486–494, Sep 2000.

[105] B. C. Bryksa, P. Bhaumik, E. Magracheva, D. C. De Moura, M. Kurylowicz, A. Zdanov, J. R. Dutcher, A. Wlodawer, and R. Y. Yada, "Structure and mechanism of the saposin-like domain of a plant aspartic protease.," *J Biol Chem*, vol. 286, pp. 28265–28275, Aug 2011.

[106] G. A. Petsko, "Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat".," *Methods Enzymol*, vol. 334, pp. 469–478, 2001.

[107] D. B. Volkin and A. M. Klibanov, "Thermal destruction processes in proteins involving cystine residues.," *J Biol Chem*, vol. 262, pp. 2945–2950, Mar 1987.

[108] H. F. Gilbert, "Thiol/disulfide exchange equilibria and disulfide bond stability.," *Methods Enzymol*, vol. 251, pp. 8–28, 1995.

[109] M. Weik, R. B. Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J. L. Sussman, "Specific chemical and structural damage to proteins produced by synchrotron radiation.," *Proc Natl Acad Sci U S A*, vol. 97, pp. 623–628, Jan 2000.

[110] R. B. Ravelli and S. M. McSweeney, "The 'fingerprint' that X-rays can leave on structures.," *Structure*, vol. 8, pp. 315–328, Mar 2000.

[111] G. Jurgens, S. Survase, O. Berezina, E. Sklavounos, J. Linnekoski, A. Kurkijarvi, M. Vakeva, A. van Heiningen, and T. Granstrom, "Butanol production from lignocellulosics.," *Biotechnol Lett*, vol. 34, pp. 1415–1434, Aug 2012.

[112] B. C. Saha, "Hemicellulose bioconversion.," *J Ind Microbiol Biotechnol*, vol. 30, pp. 279–291, May 2003.

[113] C. A. Mooney, S. D. Mansfield, M. G. touhy, and J. N. Saddler, "The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods," *Bioresour Technol*, vol. 64, no. 2, pp. 113–119, 1998.

[114] K. Shill, S. Padmanabhan, Q. Xin, J. M. Prausnitz, D. S. Clark, and H. W. Blanch, "Ionic liquid pretreatment of cellulosic biomass: enzymatic hydrolysis and ionic liquid recycle.," *Biotechnol Bioeng*, vol. 108, pp. 511–520, Mar 2011.

[115] C. Li, B. Knierim, C. Manisseri, R. Arora, H. V. Scheller, M. Auer, K. P. Vogel, B. A. Simmons, and S. Singh, "Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification.," *Bioresour Technol*, vol. 101, pp. 4900–4906, Jul 2010.

[116] Y. Wang, M. Radosevich, D. Hayes, and N. Labb, "Compatible ionic liquid-cellulases system for hydrolysis of lignocellulosic biomass.," *Biotechnol Bioeng*, vol. 108, pp. 1042–1048, May 2011.

[117] J. M. Gladden, M. Allgaier, C. S. Miller, T. C. Hazen, J. S. VanderGheynst, P. Hugenholtz, B. A. Simmons, and S. W. Singer, "Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass.," *Appl Environ Microbiol*, vol. 77, pp. 5804–5812, Aug 2011.

[118] J. I. Park, E. J. Steen, H. Burd, S. S. Evans, A. M. Redding-Johnson, T. Batth, P. I. Benke, P. D'haeseleer, N. Sun, K. L. Sale, J. D. Keasling, T. S. Lee, C. J. Petzold, A. Mukhopadhyay, S. W. Singer, B. A. Simmons, and J. M. Gladden, "A thermophilic ionic liquid-tolerant cellulase cocktail for the production of cellulosic biofuels.," *PLoS One*, vol. 7, no. 5, p. e37010, 2012.

[119] S. Datta, B. Holmes, J. I. Park, Z. Chen, D. C. Dibble, M. Hadi, H. W. Blanch, B. A. Simmons, , and R. Sapra, "Ionic liquid tolerant hyperthermophilic cellulases for biomass pretreatment and hydrolysis," *Green Chem*, vol. 12, pp. 338–345, 2010.

[120] T. Zhang, S. Datta, J. Eichler, N. Ivanova, S. D. Axen, C. A. Kerfeld, F. Chen, N. Kyrpides, P. Hugenholtz, J.-F. Cheng, K. L. Sale, B. Simmons, and E. Rubin, "Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance," *Green Chem*, vol. 13, pp. 2083–2090, 2011.

[121] M.-C. Lo, A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas, and G. Ellestad, "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery.," *Anal Biochem*, vol. 332, pp. 153–159, Sep 2004.

[122] J. J. Lavinder, S. B. Hari, B. J. Sullivan, and T. J. Magliery, "High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering.," *J Am Chem Soc*, vol. 131, pp. 3794–3795, Mar 2009.

[123] Q. Qing, B. Yang, and C. E. Wyman, "Impact of surfactants on pretreatment of corn stover.," *Bioresour Technol*, vol. 101, pp. 5941–5951, Aug 2010.

[124] T. Eriksson, J. Borjesson, and F. Tjerneld, "Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose," *Enzyme Microb Tech*, vol. 31, no. 3, pp. 353–364, 2002.

[125] B. Yang and C. E. Wyman, "BSA treatment to enhance enzymatic hydrolysis of cellulose in lignin containing substrates.," *Biotechnol Bioeng*, vol. 94, pp. 611–617, Jul 2006.

[126] S. Brethauer, M. H. Studer, B. Yang, and C. E. Wyman, "The effect of bovine serum albumin on batch and continuous enzymatic cellulose hydrolysis mixed by stirring or shaking.," *Bioresour Technol*, vol. 102, pp. 6295–6298, May 2011.

[127] F. H. Arnold and A. A. Volkov, "Directed evolution of biocatalysts.," *Curr Opin Chem Biol*, vol. 3, pp. 54–59, Feb 1999.

[128] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, "Protein stability promotes evolvability.," *Proc Natl Acad Sci U S A*, vol. 103, pp. 5869–5874, Apr 2006.

[129] S. Bershtein and D. S. Tawfik, "Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts.," *Mol Biol Evol*, vol. 25, pp. 2311–2318, Nov 2008.

[130] M. Goldsmith and D. S. Tawfik, "Directed enzyme evolution: beyond the low-hanging fruit.," *Curr Opin Struct Biol*, vol. 22, pp. 406–412, Aug 2012.

[131] S. Lutz, "Beyond directed evolution–semi-rational protein engineering and design.," *Curr Opin Biotechnol*, vol. 21, pp. 734–743, Dec 2010.

[132] B. Steipe, B. Schiller, A. Pluckthun, and S. Steinbacher, "Sequence statistics reliably predict stabilizing mutations in a protein domain.," *J Mol Biol*, vol. 240, pp. 188–192, Jul 1994.

[133] M. Lehmann, L. Pasamontes, S. F. Lassen, and M. Wyss, "The consensus concept for thermostability engineering of proteins.," *Biochim Biophys Acta*, vol. 1543, pp. 408–415, Dec 2000.

[134] M. Lehmann and M. Wyss, "Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution.," *Curr Opin Biotechnol*, vol. 12, pp. 371–375, Aug 2001.

[135] K. Watanabe, T. Masuda, H. Ohashi, H. Mihara, and Y. Suzuki, "Multiple proline substitutions cumulatively thermostabilize *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase. irrefragable proof supporting the proline rule.," *Eur J Biochem*, vol. 226, pp. 277–283, Dec 1994.

[136] M. T. Reetz, J. D. Carballeira, and A. Vogel, "Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability.," *Angew Chem Int Ed Engl*, vol. 45, pp. 7745–7751, Nov 2006.

[137] A. Bornot, C. Etchebest, and A. G. de Brevern, "Predicting protein flexibility through the prediction of local structures.," *Proteins*, vol. 79, pp. 839–852, Mar 2011.

[138] E. G. Hutchinson and J. M. Thornton, "PROMOTIF–a program to identify and analyze structural motifs in proteins.," *Protein Sci*, vol. 5, pp. 212–220, Feb 1996.

[139] R. A. Laskowski, "PDBsum: summaries and analyses of PDB structures.," *Nucleic Acids Res*, vol. 29, pp. 221–222, Jan 2001.

[140] M. W. MacArthur and J. M. Thornton, "Influence of proline residues on protein conformation.," *J Mol Biol*, vol. 218, pp. 397–412, Mar 1991.

[141] R. S. Prajapati, M. Das, S. Sreeramulu, M. Sirajuddin, S. Srinivasan, V. Krishnamurthy, R. Ranjani, C. Ramakrishnan, and R. Varadarajan, "Thermodynamic effects of proline introduction on protein stability.," *Proteins*, vol. 66, pp. 480–491, Feb 2007.

[142] E. G. Hutchinson and J. M. Thornton, "A revised set of potentials for beta-turn formation in proteins.," *Protein Sci*, vol. 3, pp. 2207–2216, Dec 1994.

[143] B. W. Matthews, H. Nicholson, and W. J. Becktel, "Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding.," *Proc Natl Acad Sci U S A*, vol. 84, pp. 6663–6667, Oct 1987.

[144] E. C. Ohage, W. Graml, M. M. Walter, S. Steinbacher, and B. Steipe, "Beta-turn propensities as paradigms for the analysis of structural motifs to engineer protein stability.," *Protein Sci*, vol. 6, pp. 233–241, Jan 1997.

[145] K. Bajaj, M. S. Madhusudhan, B. V. Adkar, P. Chakrabarti, C. Ramakrishnan, A. Sali, and R. Varadarajan, "Stereochemical criteria for prediction of the effects of proline mutations on protein stability.," *PLoS Comput Biol*, vol. 3, p. e241, Dec 2007.

[146] S. R. Trevino, S. Schaefer, J. M. Scholtz, and C. N. Pace, "Increasing protein conformational stability by optimizing beta-turn sequence.," *J Mol Biol*, vol. 373, pp. 211–218, Oct 2007.

[147] H. Fu, G. R. Grimsley, A. Razvi, J. M. Scholtz, and C. N. Pace, "Increasing protein stability by improving beta-turns.," *Proteins*, vol. 77, pp. 491–498, Nov 2009.

[148] R. Guerois, J. E. Nielsen, and L. Serrano, "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.," *J Mol Biol*, vol. 320, pp. 369–387, Jul 2002.

[149] K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, "Predicting DNA duplex stability from the base sequence.," *Proc Natl Acad Sci U S A*, vol. 83, pp. 3746–3750, Jun 1986.

[150] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, *Protein identification and analysis tools on the ExPASy server*, pp. 571–607. Humana Press Inc., Totowa, NJ, 2005.

[151] N. Nelson, "A photometric adaptation of the Somogyi method for the determination of glucose," *J Biol Chem*, vol. 153, pp. 375–80, 1944.

[152] M. Somogyi, "Notes on sugar determination by Somogyi," *J Biol Chem*, vol. 195, pp. 19–23, 1951.

[153] J. T. Park and M. J. Johnson, "A submicrodetermination of glucose.," *J Biol Chem*, vol. 181, pp. 149–151, Nov 1949.

[154] G. L. Miller, "Use of dinitrosalicyclic acid reagent for determination of reducing sugar," *Anal Chem*, vol. 31, no. 3, pp. 426–428, 1959.

Index